

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:50:18 ; Search time 65 Seconds
(without alignments)
5229.302 Million cell updates/sec

Title: X69111

Perfect score: 7544

Sequence: 1 GARCCTGGGTGCTGCAGGA.....TTAACTTTTATATAAAGTT 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3504.5	46.5	3907	5	ABG70822 Mouse myo
2	3504.5	46.5	3907	6	ABG74190 Mouse myo
3	3447.5	45.7	2211	6	ABG72959 Novel hum
4	3322	44.0	7285	6	ABJ38280 PAMG21-RA
5	3215.5	42.6	1879	4	AU07343 l-aminocyc
6	3173.5	42.1	1811	6	AAS37007 Micronono
7	3105	41.2	2601	6	ADA15723 C. elegan
8	3103.5	41.1	1329	5	AU91279 Human NOV
9	3101.5	41.1	1743	6	ABU88255 Novel hum
10	3101.5	41.1	1743	6	ABU90134 Novel hum
11	3101.5	41.1	1743	6	ABU96436 Novel hum
12	3101.5	41.1	1743	6	ABU99045 Novel hum
13	3101.5	41.1	1743	6	ABU98260 Novel hum
14	3101.5	41.1	1743	6	ABU91966 Novel hum
15	3101.5	41.1	1743	6	ABU85270 Novel hum
16	3101.5	41.1	1743	6	ABU00409 Novel hum
17	3101.5	41.1	1743	6	ABU88960 Novel hum
18	3101.5	41.1	1743	6	ABU06456 Novel hum
19	3101.5	41.1	1743	6	ABU95516 Novel hum
20	3101.5	41.1	1743	6	ABU95206 Novel hum
21	3101.5	41.1	1743	6	ABU90754 Novel hum
22	3101.5	41.1	1743	6	ABU93916 Novel hum
23	3101.5	41.1	1743	6	ABU86190 Novel hum
24	3101.5	41.1	1743	6	ABU82845 Novel hum
25	3101.5	41.1	1743	6	ABU07906 Novel hum

26 3101.5 41.1 1743 6 ABU94226 Novel hum

27 3101.5 41.1 1743 6 ABO00099 Novel hum

28 3101.5 41.1 1743 6 ABU87110 Novel hum

29 3101.5 41.1 1743 6 ABU91351 Novel hum

30 3101.5 41.1 1743 6 ABU90444 Novel hum

31 3101.5 41.1 1743 6 ABU97035 Novel hum

32 3101.5 41.1 1743 6 ABO05231 Novel hum

33 3097.5 41.1 2508 6 ADA15721 C. elegan

34 3097.5 41.1 2544 6 ADA15717 C. elegan

35 3082.5 40.9 1300 6 ABU88254 Novel hum

36 3082.5 40.9 1300 6 ABU90133 Novel hum

37 3082.5 40.9 1300 6 ABU96435 Novel hum

38 3082.5 40.9 1300 6 ABU95044 Novel hum

39 3082.5 40.9 1300 6 ABU98259 Novel hum

40 3082.5 40.9 1300 6 ABU91965 Novel hum

41 3082.5 40.9 1300 6 ABU85269 Novel hum

42 3082.5 40.9 1300 6 ABO00408 Novel hum

43 3082.5 40.9 1300 6 ABU88959 Novel hum

44 3082.5 40.9 1300 6 ABO06455 Novel hum

45 3082.5 40.9 1300 6 ABU95515 Novel hum

ALIGNMENTS

RESULT 1

ABG70822

ID ABG70822 standard; protein; 3907 AA.

AC ABG70822;

DT 16-DEC-2002 (first entry)

DE Mouse myocardin associated protein #1.

KW Mouse: myocardin; myocardial infarction; cardiomyocyte;

KW Post-mitotic cell; differentiation; therapeutic; gene therapy;

KW heart disease; cardiomyopathy; mortality; heart failure; hypertension.

OS Mus sp.

PN WC200260946-A2.

PD 08-AUG-2002.

PF 21-DEC-2001; 2001WO-US050606.

PR 21-DEC-2000; 2000US-0257716P.

(TEXA) UNIV TEXAS SYSTEM.

Olson EN, Wang D;

WPI; 2002-732693/79.

New myocardin polypeptides and polynucleotides, useful for respecifying non-cardiac cells, stimulating cardiac tissue regeneration, and for treating cardiovascular disorders, such as myocardial infarction and hypertension.

Disclosure; Page 146-159; 175pp; English.

The invention discloses an isolated polynucleotide encoding a myocardin polypeptide. Myocardial infarction results in the loss of cardiomyocytes, which are post-mitotic cells and generally do not regenerate after birth. Transplanting fetal cardiomyocytes has limitations so identifying new regulators of cardiomyocyte growth and differentiation is an important goal in the search for therapeutics to treat myocardial tissue damage. The polynucleotides, polypeptides and methods of the invention can be used to modulate the phenotype of a non-cardiomyocyte cell to include one or more phenotypic functions of a cardiomyocyte cell, to generate a cardiomyocyte, which comprises introducing into a cardiac fibroblast the myocardin polynucleotide and a promoter which is active in the

fibroblast, stimulate cardiac tissue regeneration which comprises inhibiting the function of myocardin in a post-mitotic cardiomyocyte and for screening for a modulator of myocardin expression. The nucleic acid can also be used in gene therapy to treat a heart disease, including cardiomyopathy, comprising administering a polynucleotide encoding a myocardin peptide or protein or an antisense nucleic acid. The polynucleotide and polypeptide can also be used for decreasing mortality in a subject with heart failure, comprising inhibiting the function of myocardin in post-mitotic cardiomyocytes, increasing the level of myocardin in fibroblasts to generate cardiomyocytes, inhibiting the function of myocardin in post-mitotic cardiomyocytes or increasing the level of myocardin in fibroblasts to generate cardiomyocytes in the subject. The compositions and methods of the present invention are useful for respecifying non-cardiac cells, stimulating cardiac tissue regeneration, for treating cardiovascular disorders, such as myocardial infarction and hypertension, and for screening compounds for various abilities to interact and/or affect myocardin expression or function. The sequence presented is the mouse myocardin associated protein, #1. NOTE: This sequence is presented as a 3-letter coded protein sequence in the specification, but when changed into the single letter code, forms a DNA sequence identical to that given in AH55224. It, therefore, appears to be the DNA sequence encoding human myocardin 2

Query Match	46.5%; Score 3504.5; DB 5; Length 3907;
Best Local Similarity	36.8%; Pred. No. 4.9e-148;
Matches 948; Conservative	0; Mismatches 347; Indels 1111; Gaps 134;
QY	1 GATCTGGGGTGCTGCC-----AGGAAAAGC-----AAATTCTG--GAAGTTAATGGT 46
DB	754 GGACAGGGGGGCACC CCCCATGTGACTCATCTTAGCCAGAATCCTTCGACGACGACGT 813
QY	47 TTT---GAG---TGATTTTTAA-----AT 64
DB	814 CTTCTCCAGCTGCAGATCTCAACAGCAGCMGACGACGACCAACTACCGGCCAT 873
QY	65 CCTTCTGGC-----GGAGAGGCC----- 84
DB	874 CTGCTGTCCC CGCAAAGT CAGCAGCGCGAGCCCTGGGAAGCAGCGGACCCCCCAGT 933
QY	85 ---GCTCTCC---CCGTATCACG-----GCT 106
DB	934 ACGCAGCTCTCCACTACCAATAGCAGCTCCAGCTCGGGCGCCCTGGGCGCTGTGGGCT 993
QY	107 -----TCCTCAT--CTTTGAATCGCG-GCTCG-GCGTCTTC----- 141
DB	994 GGCAGCTCAGAACAGCAGACCTCCTGACTGGCAAGCGGGAGCCCTGCGCGCAACCTGGA 1053
QY	142 -----GGCGTCAGACC--AGCGGAG-GAA-----GCCTGTTT--G 172
DB	1054 CGACATGAAGTGTGCAGAGCTGAAGCAGAGGCTGAAGTTGCATCACTGCTGTCTCGGG 1113
QY	173 CAATTTAAGCGGGTG--TCAAGCC-----CAGGGC-----GGC 206
DB	1114 CACCAAACTGAGCTGATTGAGGGCTTTCGAGCCTATCAAGCAAAATCAGCCCTGTGCC 1173
QY	207 GGGG----CGGGCGCAGCGCG--GCC---ATTTTGAATAAG-AGCGC----- 246
DB	1174 AGGAGCCCCAAGGCCCTCCGCCACCTCTATCTCTGCACAAGGCTGCGCAGGTGGTGGT 1233
QY	247 TGCTTTCAGG---CAGGCTCTAATAGTCAAC-GCC-----GGGGCGAGCGT-GGCG- 293
DB	1234 AGCCTTCCACGCGCCCGCTGAGCAGCGGGCCAGCCCTGGTGGCAGCAGSCCTGGCTCC 1293
QY	294 ----GCGTTGCAGGTCACTAG-----CGGA 316
DB	1294 AGCTAGTGTGTGTGCCACGGTGGCCAGTGGGGTGTGAAGTTTGGCAGCACGGG 1353
QY	317 CT-----TCT-----TTGGTTTCTTTCT----- 336
DB	1354 CTCACGCCCCCGTGTCTCCACCCCTCCGAGGCGCTCACTGTCTCAGCAGCGCGATGA 1413

337	QY	-----	-----	CTTTGGGG	-----	CACTCTGT	-----	GAC	355
1414	DB	AAACTCCACCCCGGGGACACCTTTGGT	GAGT	GAGT	GATGAT	CATC	ACTCTGTG	CGCAGCTGAC	1473
356	QY	-----	-----	TCACTCCCCAGCA	-----	TGAAGCG	-----	376	
1474	DB	CCTGACGGCTCGCCACTG	CAGAT	CCTCGT	GAA	GAGGAGG	GGCCCCCGGGCGGGTCTTG	1533	
377	QY	-----	-----	CTGAGCCC	-----	GCTGCGCG	-----	GTTGC	397
1534	DB	TTGCTTGAGCCCTTGGGGGGGGCGGAG	CTAG	AGGGCGCG	CAAG	GACACAGAT	TGCTGCA	1593	
398	QY	-----	-----	TACGAGCGGGTG	-----	TGCTGC	-----	CTGTG	420
1594	DB	GGAGAAAGACAAGCAGAT	CGAGCG	CTGA	CGCGCAT	CTCGG	CAGAGCAGCAGCTGGT	1653	
421	QY	-----	-----	GGAAAC	-----	GCATCTG	-----	GCCAT	438
1654	DB	GGACGGCTCAAGT	GTGAG	CTGAG	CAGGAG	AGCGCC	CAGCAGCCCGCCCCCCC	1713	
439	QY	CGCC	-----	CGGGGCC	-----	GAGGAGAGGGCCGGCAGCT	GAGGAGCC	GTTGAGCTT	486
1714	DB	CGCCCCCTTGG	CACCCCGTGA	CAGAGAG	CAACGCTT	CTCCAGCT	TGCCAGCTGACGACCA	1773	
487	QY	G	-----	CTGGACGACATGAACACTGT	CTAC	CTCCGGCT	CGCG	-----	523
1774	DB	GCAGCCCTGGGCCCCGCTCAGCCAT	TTC	CAACCC	CAGCCT	GGCGCCCC	CAGCCACCA	1833	
524	QY	-----	-----	GAACTG	GTACCCGAGTCC	-----	GAGAGGCA	-----	551
1834	DB	CATAGACCTTGTGCT	TGCCCCGGGGCC	CCCGT	CGTGTGT	GTTGA	GACAGGAAGCTT	1893	
552	QY	-----	-----	CTGAG	-----	CTTAG	-----	561	
1894	DB	GCAGCTGACCCGACCGGCT	CCCCCGCC	CGAGTTG	TTCTG	GGGGCT	CAGGGCCCCAG	1953	
562	QY	CC	-----	AGGTGG	-----	AAATCT	CACAGCGCTCAT	CGACTACA	596
1954	DB	CCTCATCAAGGGGTTGC	ACTCCACC	CTCAT	CACCGA	CTCC	CAGGACCC	CACCTTG	2012
597	QY	TTCTC	-----	GACC	-----	TGCAGGTAGT	CTCTGGCC	-----	627
2013	DB	TCCTCACCGTGACCA	TAAAGATG	CAGACAG	CCCTGG	CTGTCC	AGTGGGAGCCCCAGC	2072	
628	QY	AGCC	-----	CTGG	-----	ACCCCT	-----	GATGG	648
2073	DB	AGCCCTCGTCCC	AGCTGGCT	TTCC	AGCGCT	GTGCCCCCT	GTGCC	CAGATGGA	2132
649	QY	-----	-----	CCCCAC	-----	CTTC	-----	CCATCCAGAC	669
2133	DB	ACCACTGACGCCCT	CTTTTGG	AGCCCC	CTCTG	CTGA	AGAGGAACCA	CTGGC	2191
670	QY	-----	-----	AGCC	-----	GAGCTCGT	-----	CGGAACTTG	701
2192	DB	TATGAGAAACCATG	AGCCAG	CAGCCCCAA	CAGCAG	GAATGTTCT	CAAGCCAGCAG	2251	
702	QY	-----	-----	ACGAC	-----	AAGAGGCTTTTGG	CCACTGACT	-----	729
2252	DB	ATGACGACCTGTT	GCATTT	CTCAT	TT	CAGAGCGGAGAA	TTTCAG	CAGATTTCA	2311
730	QY	CGGCGGTGT	CCTGACA	-----	-----	CCT	-----	CCAGAACGAG	764
2312	DB	CCGCCAT	CCCTGCC	AGGAGGAG	AGC	ATCCCG	GAAGAC	AGTCTGTGGT	2370
765	QY	GGC	-----	CCC	-----	CGTTCTG	-----	CTGG	794
2371	DB	GGCAGCAGCCAT	CACCTT	CTGT	GAGCT	CCCCCAG	CTGCC	CACTCT	2430
795	QY	CTCTCTG	CC	-----	GGAGCG	AGGATGG	-----	GCCCACT	835
2431	DB	ACCTCCCTCC	TGGAGCGCT	GGAGCA	TTCT	GTGAG	CAGCAGCA	CGGGGTG	2490
836	QY	-----	-----	TGG	-----	CCCTGCCAC	TTGAC	TTACCAAAATC	854

Db 2491 GACCAAGTGGGATGAGCGGCGCAGAGCCCTTTCTCTTATTTACGACTTCTTACCCAGAT 2550
 Qy 865 -----CCTTCTGGAGACTAAACCTGGTCTCAGAGC-----GAA--GG 902
 Db 2551 GGTGAGCAGCACTGCATCCTGGA--CCACCCCGCTCACCCTACGACACCTCGGAATTGC 2609
 Qy 903 ACTGTGAACCTGTAGCTGAGAGGCA-----GAGCTAGCT--CTGGCCACC----- 947
 Db 2610 ACTTGTCTGAGCGCAGCAGACATGCGGCTGAGACTGCTGTATGCGCCACTTGACACA 2669
 Qy 948 -----AGCTGG--GC--GAGCTCA-----CCC--TGCT-----CCACCC----- 975
 Db 2670 GCATGAGCTGGCTGGAGCTGTCTGAGTGTGTCTGAGCTGAGCTAGCCCTCCAGCA 2729
 Qy 976 CCAC--CCCCAAG--TTCT--AAGTCTTTTCA--GGTGTGAGTGTGGAAGAGT 1025
 Db 2730 CCACAGCCCCAGCCTTCTTCCAGACTTCTCTGATGGCCATGATTTGAGCTGCACT 2789
 Qy 1026 GG-----CTGCT-----CTC--CAAACT--ATGCCAAGGCGGCGCAGAGCTGG----- 1065
 Db 2790 GGGATTCTCTGTGTAGCTCTCTGCTCAAGACGCGGTGTGGGAAGGGGCTGGAGCCAGG 2849
 Qy 1066 --TCTTC-----TGG--TCTCTTGGAGA-----AAGGTTCTG-----TTG-- 1097
 Db 2850 GATCTCCATGGTGGCTCTCTGCTGATTCGGGCTCTCCATGTTGTGATCTTGA 2909
 Qy 1098 -----CCTGATTTAT-----GAACTCTATATAGAGTA-----TAT 1129
 Db 2910 CAATCACAGCCCTGCTTTTCTCCCTTCCCTGCGAGGCTAGAACAGAGAGCCCTTACTCC 2969
 Qy 1130 AGTTTGTGACCTTTTATACAGG---AAGGTGACTTTCTGTAAACAATGCG---ATGTATA 1183
 Db 2970 TGGTTGAGTGC-----ACCAGGCGAGGAGGAGCAGCTGTTCAGAGCAGCCCTGGCTC 3025
 Qy 1184 TTAAACT-----TTTT--ATPAAAGTT 1203
 Db 3026 TCACGCTGGGTTTGGACACACGGT 3051

RESULT 2
 ABG74190
 ID ABG74190 standard; protein; 3907 AA.
 AC ABG74190;
 XX
 DT 29-APR-2003 (first entry)
 XX
 DE Mouse myocardin associated protein.
 XX
 KW Mouse; myocardin; cardiomyocyte growth; ANF induction;
 KW cardiomyocyte differentiation; sarcomere assembly induction; ANF;
 KW cardiac hypertrophy; atrial natriuretic factor; fibroblast modulation;
 KW non-cardiomyocyte cell modulation; cardiomyocyte; heart disease;
 KW cardiomyopathy; myocardial infarction; hypertension; gene therapy;
 KW cardiac-specific transcription factor.
 XX
 OS Mus musculus.
 XX
 PN US2002164735-A1.
 XX
 XX 07-NOV-2002.
 XX
 XX 21-DEC-2001; 2001US-00029217.
 XX
 XX 21-DEC-2000; 2000US-0257761P.
 XX
 PA (OLSON/) OLSON E N.
 PA (WANG/) WANG D.
 XX
 PI Olson EN, Wang D;
 XX
 XX WPI; 2003-247258/24.

XX Novel isolated polynucleotide encoding human or murine myocardin 1
 PT polypeptide, useful for modulating phenotype of non-cardiomyocyte cell
 PT e.g., fibroblast, to include phenotypic functions of cardiomyocyte cell.
 XX
 PS Disclosure; Page 51-61; 104pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding myocardin
 CC polypeptide. The effects of myocardin in growth and/or all
 CC differentiation of cardiomyocytes was assessed by overexpressing
 CC myocardin in cardiomyocytes using adenoviral delivering system. The
 CC results showed that overexpression of myocardin in neonatal
 CC cardiomyocytes induced assembly of sarcomeres and expression of atrial
 CC natriuretic factor, ANF, markers of cardiac hypertrophy. An expression
 CC cassette containing the polynucleotide operably linked to a regulatory
 CC sequence is useful for modulating the phenotype of a non-cardiomyocyte
 CC cell e.g. fibroblast to include one or more phenotypic functions of a
 CC cardiomyocyte cell. The expression cassette is useful for generating a
 CC cardiomyocyte which involves introducing into a cardiac fibroblast the
 CC expression cassette comprising the polynucleotide and a promoter active
 CC in the fibroblast, where the promoter directs the expression of the
 CC polypeptide. The expression cassette further comprises a second
 CC polynucleotide encoding GATA4, under the control of a second promoter
 CC active in a cardiac fibroblast. The expression cassette further comprises
 CC a polyadenylation site and an immunological marker. An expression or
 CC cassette comprising the polynucleotide encoding myocardin protein or
 CC peptide and a promoter operable in eukaryotic cells is useful for
 CC treating a heart disease, including cardiomyopathy, such as myocardial
 CC infarction or hypertension. The present sequence represents the amino
 CC acid sequence of the mouse myocardin associated protein. Note: The
 CC protein sequence presented is not disclosed in the specification but is
 CC shown in the sequence listing
 XX
 SQ Sequence 3907 AA;

Query Match 46.5%; Score 3504.5; DB 6; Length 3907;
 Best Local Similarity 36.8%; Pred. No. 4.9e-148;
 Matches 848; Conservative 0; Mismatches 347; Indels 1111; Gaps 134;
 Qy 1 GATCTGGGGTGTGCTCC-----AGGAAAAGC-----AAATCTG--GAGGTTAATGCT 46
 Db 754 GGCAGAGGGGACCCCCCATGGACTATCTAGCCCAAGATCTGAGCAGCAGCAGCT 813
 Qy 47 TTT-----GAG-----TGATTTTAA-----AT 64
 Db 814 CTTCCTCCAGCTGCAGATCTCTCAACAGCAGCAGCAGCAGCAGCACTACCCAGGCCAT 873
 Qy 65 CTTTCTGCTGGC-----GGAGAGGCC-----84
 Db 874 CTTGCTCTCCCGCCCAAGTCAGCAGCAGGAGGCCCTGGGAGCAGCGGGACCCCCAGT 933
 Qy 85 -----GCTCTTC-----CCGTATCAGC-----GCT 106
 Db 934 AGCAGCCTCTCCACTACCAATAGCAGCTCCAGCTCGGGCGCCCTCGGGCCCTGTGGGCT 993
 Qy 107 -----TCTTCATT--CTTTGAATCCGCG--GCTCCG--CGGTCTTC----- 141
 Db 994 GGCAGCTCAGAACAGCAGCTCACTGCTGGCAGCGGGAGCCCTGCCCACTGGA 1053
 Qy 142 -----GGCTCAGACC---AGCCGAG--GAA-----GCTGTGTT---G 172
 Db 1054 CGACATGAAGTGGCAGAGCTGAGCAGGAGCTGAAGTTCGATCACTGCTGTCTCGGG 1113
 Qy 173 CAATTTAAGCGGGTGG--TGAACGCC-----CAGGGCC-----GGC 206
 Db 1114 CACCAAACTGAGCTGATTGAGCGCTTCGAGCCCTATCAGACCAATATCAGCCCTGTGCC 1173
 Qy 207 GGGGG-----CGGGCCGAGGCGG--GCC---ATTTGAATTAAG--AGCG----- 246
 Db 1174 AGGAGCCCCCAAGGCCCTGCGCCACCTCTATCTGTCAAGAGCTCGAGAGTGTGGT 1233
 Qy 247 TGCCTTCCAGG-----CAGGCTCTATAAGTGACC--GCC-----GCGGCGAGCGT--GGCG-- 293

PR 18-DEC-2000; 2000US-00738884.
 XX (PEKE) PE CORP NY.
 XX Guegler K, Beasley EM, Ketchum KA, Di Francesco V;
 XX WPI; 2003-182638/18.
 XX Novel isolated human phospholipase peptide and gene encoding the protein,
 PT useful for diagnosing, treating and preventing disorders associated with
 PT aberrant or unwanted phospholipase expression.
 XX Disclosure; Page 20-25; 57pp; English.
 XX The invention describes an isolated human phospholipase polypeptide (I),
 CC its allelic variant or orthologue. (I) and the polynucleotide encoding it
 CC (II) are useful as models for the development of human therapeutic
 CC targets, for identifying therapeutic proteins, and as targets for
 CC development of human therapeutic agents that modulate phospholipase
 CC protein activity in cells and tissues. (I) is useful in assays to
 CC determine the biological activity of the protein and levels of the
 CC protein in biological fluids, as markers for tissues in which the
 CC corresponding protein is preferentially expressed, to identify binding
 CC partners to develop a system to identify inhibitors of binding
 CC interaction, and in the treatment of disorders characterized by an
 CC absence of or unwanted expression of the protein. (I) is also useful in
 CC drug screening assays, to identify compounds that modulate phospholipase
 CC protein activity of the protein. (I) is also useful for screening a
 CC compound for its ability to stimulate or inhibit interaction between the
 CC phospholipase protein and a molecule that normally interacts with the
 CC phospholipase protein, in pharmacogenomic analysis, and for treating a
 CC disorder characterized by an absence of, in appropriate or unwanted
 CC expression of the human phospholipase protein. (II) is useful for
 CC constructing recombinant vectors, host cells and transgenic animals, for
 CC expressing antigenic portions of the human phospholipase proteins, for
 CC designing ribozymes, for monitoring the effectiveness of modulating
 CC compounds on the expression or activity of a human phospholipase protein
 CC gene, as an antisense construct to control human phospholipase protein
 CC gene expression in cells, tissues and organisms, and for gene therapy.
 CC This is the amino acid sequence of a novel human phospholipase C
 CC subfamily phospholipase associated protein. Note: This sequence appears
 CC in the sequence listing as SEQ ID number 1 but is not described in the
 CC specification
 XX Sequence 2211 AA;
 QY Query Match 45.7%; Score 3447.5; DB 6; Length 2211;
 DB Best Local Similarity 36.9%; Pred. No. 1.1e-145;
 DB Matches 809; Conservative 0; Mismatches 355; Indels 1027; Gaps 124;
 QY 4 CTGGGGTCTGCCAGGAAAGCAATCTGGAGTTAA--TCGTTTT--GAGTGAATTT 59
 DB 37 CTGGGGGCTCCC--GGCTCCGAGATCGCTGGCGACGTGGCAGAGAGGGCTGT 94
 QY 60 TAAATCCTTGTGGCGAGAGGCC-----CGCTCTCCCG----- 95
 DB 95 ACCGGC--TGCAGAGGA--CGGCTGAGCGTGTGTTCCAGCGCGGCMTCGCGTGGC 151
 QY 96 -GTATCAGCGTCTCTCAT-----TCTTTGA-----ATCCGC-----TCGC 133
 DB 152 CATCGAGACATCTTCTGTCAGCATCGAGCGGTGCGAGGCGCCACCATCG 211
 QY 134 C-GGTCTTCGGG--TCAG-----ACC-----AGCCGAG--GAAGCCT---GTTTCAATTT 178
 DB 212 AGGCGCTTGGCGGCTTCTGGGGTGCCTTCGCGCCAGCGGTGCTCACCATCGCTTCA 271
 QY 179 AAG-----CGGG--CTGTGAACGCCAGGGC-----CGGC---GGG 209
 DB 272 AGGCGCGCGCAAGACCTTGGACCTTGGCGGCGCCACCGCTGAGGAGCGCGCTGG 331
 QY 210 GGCGGG-----CCGAG--CGG---GGCCATTTTGAATAA-----AGAGGGGTG 248
 DB 332 TGGCGGTCTGTACCAAGCTCCGCGCGCTGGAGCCATGAGCCAGCGCGAGGGGTAG 391
 QY 249 -----CCT-----TCCAGGCGAGGCTCTATAAGTGACCGC----- 277
 DB 392 ACCACTGGATCCACTCCTATCTGCACCGGGTGTCTCCACAGGACGAGAGTGTAGCT 451
 QY 278 -CGGCGAGAGCTGCGCGCTTG-----CAGGTCACTGTAGC---GGACTTCTTTTG--GTTT 329
 DB 452 TCAAGGAGATCAAGAGCTGTCTGAGATGGTCAACGCTGACATGAACACATGTACGCT 511
 QY 330 TCTTTCTCTTTTGGGCG-----ACCTCTGTGACTCACTCC--CCAGCATGAAGCGCTGAGC 382
 DB 512 ACTCTCTCTTCAAGAGTGTGACCACT---CCACACAGCAGCGTCTAGAGGGGCTGAGA 568
 QY 393 CCGG-----TGC-GCGGCTGCT----- 398
 DB 569 TCGAGGAGTTCCTGGGGGCTGTCTCAAGCGCGCGAGCTGGAGGAGATCTTCCATCAGT 628
 QY 399 -----ACGAGG--CGGTGTGCTG-----CCTGTCCGAAAG-- 426
 DB 629 ACTCGGCGAGGACCGCTGTGTGAGTGGCCCTGAGCTGCTGAGCTTCTGGAGGACCGAG 688
 QY 427 -----CAGTCTGGCATCGCCCG-----GGCC---GAG-----G 453
 DB 689 GCGAGGAGGCGCCACACTGGCGCGCGCCAGCAGCTCATTCAGACCTATGAGCTCAACG 748
 QY 454 GAAGGCGCGCGCAGC--TGA---GGAGCGGCTG---AGCT----- 485
 DB 749 AGACAGCAAGCAGCATGAGCTGATGACACTGGATGGCTTCATGATGTACCTGTGTGTCG 808
 QY 486 -----TGC--TGCAC-----GACATGAAC----- 503
 DB 809 CGGAGGAGGACTCCCTTGGACAACACCCACAGTGTGTGTTCCAGGACATGAACAGCCCC 869
 QY 504 -----ACTGTCTACTC--CCGCTCGCGGA-----ACTGGTACCC-----GGA 538
 DB 869 TTGCCACACTTCT 928
 QY 539 GTCCCGAGGAGGACTCAG-----CTTAG--CCAG-----GTGGA 571
 DB 929 GGCC--AGCAGCAGCGAGGCTTGTAGGCGCTTGTGCCAGGATGCCGTGCTGGAG 987
 QY 572 AT--CCTAC-----AGCGGCTCTATCGA-----C 592
 DB 988 CTGGAAGTGTGGAGGGGCGAGGAGGCGCCCTCATCTATCATGCGCATCCCTCAC 1047
 QY 593 TAC---ATTCTGAGCTGAGGTAGTCTGAGCGAGC-----CAG--CC 631
 DB 1048 TCAAGATTTCTTTCGGGAGCTGGCCCAAGCCCGTGGCGGACCATGCTTCACGCTGTC 1107
 QY 632 CTGGAGC-----CCCTG-----ATG 647
 DB 1108 CTTTACCTGTCTATCT 1167
 QY 648 GCGC--CCACCT---TCCCATCCAGCAGCGAGCT---CGCT----- 682
 DB 1168 GCGCGCAGCTCTGACCATCTCTGGGAGCATGTGTGTGACACAGGCGCTGGATCCCCA 1227
 QY 683 ---CC--GGAATTTGTCTATCTCC--AACGACAAAGAGAGC---TTTTG----- 720
 DB 1228 AATCCGAGGAGC--TGCATCCCGAGCAGCTGAAGGCGCGGCTCTGTGTGAAGGAAA 1286
 QY 721 -----CCACTGACTC-----GGCGGTGTCTCTGAC----- 744
 DB 1287 GAAGTGTCCCGGTG--CTCGAGCAGGAGTGCCTGGGCTCTGTCTGATCGGAGAGGGGG 1345
 QY 745 -----AC-----TCCAGAACG----- 756
 DB 1346 AGGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1405
 QY 757 -----CAGGTGTGTGC--GCC-----GTTCTGCTGGGACCCCG-----GGA 792
 DB 1406 AGCAGATCTCCCGAGGCTGTGGCCCTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1465

QY 793 ACCT-----CT-----CCTGCGG----- 806
 Db 1466 CCTGACCTGCCCCCAAGCCCAACACCTGCGAGGTGAGTCCCTCAGCGAGCGCA 1525
 QY 807 AAGC-----CGACGCGAGGCA-----TG-----GGC-----CCCACTTC 837
 Db 1526 AAGCCCAAGAAACCTCATTCGGGAGGCGAGGGAACAGCTTTGTTCAGGCACAAATCCCGCCAGC 1585
 QY 838 --GCC-----TGCCCACT-----TGACTTCA--CCAAATCCCTTCTCTGAGACT 878
 Db 1586 TGACCCGCGGTATACCGCTGGGCTGCGGATGACTCAGCCACTACAGTCC----- 1637
 QY 879 AAACCTGGTCTCAGAGCGAGGAGGACTGTGAATTTGTAGCC--TGAGAGCGCA----- 931
 Db 1638 ---CCAGGAGATGTGAAC--TGGGGCTGTGAGTGGTGGCTTGAATTCAGACGCCAG 1693
 QY 932 GCTA-GCTCTGG--CCACCA-GCTGGGCG-----ACGTC- 961
 Db 1694 GCTACGAGATGAGACTCAATGCGGCGCTTCTTAGTCAATGAGGCGAGTGGCTACGTCC 1753
 QY 962 ---ACCTGCT--CC-----CAGCC--CAC-----CCCCAAGTCTAAGTCTTT 999
 Db 1754 TAAACCTGCTGCTCGGCGAATCTGACTGACCTTTGACCCCGAGTACCCAGGACCTC 1813
 QY 1000 TTAG-----AGCGTGGAGGTG-----TGGAAG--G 1022
 Db 1814 CCAGAACCACTCTCAGCATCCAGGTGCTGACTGCACAGCAGCTGCCCAAGCTGAATGCG 1873
 QY 1023 AGTGGCTGCTC-----CAGAGTCTGCTCAACATGCTTCAACCCCGCTGGGGC 1933
 Db 1874 AAGAGCACTCTCATCTGTGACCCCTGCTGCGCATTTGAGATCATGCGGTGCCCCGAG 1933
 QY 1039 ACTATGCGAAG-----CTGA-TTTATG-----AACTCT--ATAATAG-AG 1124
 Db 2052 AGATTATGAGCCACCTCCGCCAATGACTTTGTGGGCGAGTTTACACTGCTCTTAGCAG 2111
 QY 1125 TATATA-----GGTT-----TTGTACCTTTTATAC--AGGAAGTGAATCTGTGAACA 1171
 Db 2112 CCTAAGCAAGGTACCGGCACATACACTGCTTTCCAAAGACGGGGCTCACTGTACC 2171
 QY 1172 ATCGGATGATATTAACCTTTTATATAAGT 1202
 Db 2172 A-GCCAG-----CTCTCATCCAAAT 2192

RESULT 4

ABJ38280
 ID ABJ38280 standard; protein; 7285 AA.

AC ABJ38280;

DT 12-JUN-2003 (first entry)

XX PAMG21-RANK-Fc vector protein SEQ ID No 28.

XX TALL-1-binding protein; TALL-1; B-cell-mediated autoimmune disease;
 KW systemic lupus erythematosus; B-cell-mediated cancer; lymphoma;
 KW inflammation; rheumatoid arthritis; acute pancreatitis; atherosclerosis;
 KW Alzheimer's disease; asthma; cachexia; cirrhosis; diabetes; osteoporosis;
 KW glomerulonephritis; Hashimoto's thyroiditis; ischemic injury; psoriasis;
 KW multiple myeloma; multiple sclerosis; Parkinson's disease; vasculitis;
 KW gene therapy; ds.

OS Unidentified.

PN WO200292620-A2.
 XX 21-NOV-2002.
 PF 13-MAY-2002; 2002WO-US015273.
 XX 11-MAY-2001; 2001US-0290196P.
 XX (AMGE-) AMGEN INC.
 XX Min H, Hsu H;
 XX WPI; 2003-156719/15.
 XX New TALL-1-binding polypeptide, useful for modulating the activity of
 PT TALL-1 and in treating, preventing or diagnosing a B-cell-mediated
 PT autoimmune diseases, cancers or lymphomas.
 XX Disclosure; Fig 4; 236pp; English.
 CC The invention relates to a novel TALL-1-binding polypeptide comprising a
 CC defined sequence in the specification. The composition is useful in
 CC modulating the activity of TALL-1, and in treating, preventing,
 CC ameliorating, diagnosing or prognosing a B-cell-mediated autoimmune
 CC disease (e.g. systemic lupus erythematosus) or B-cell-mediated cancer or
 CC lymphoma. The composition may also be used in treating inflammations
 CC (e.g. rheumatoid arthritis), acute pancreatitis, Alzheimer's disease,
 CC asthma, atherosclerosis, cachexia, cirrhosis, diabetes,
 CC glomerulonephritis, Hashimoto's thyroiditis, ischaemic injury, multiple
 CC myeloma, multiple sclerosis, osteoporosis, Parkinson's disease, psoriasis
 CC and vasculitis. Disorders may be treated with the novel composition using
 CC gene therapy. This polynucleotide sequence represents a TALL-1 related
 CC DNA sequence of the invention
 XX
 SQ Sequence 7285 AA;

Query Match 44.0%; Score 3322; DB 6; Length 7285;

Best Local Similarity 37.8%; Pred.No. 1e-139;

Matches 787; Conservative 0; Mismatches 396; Indels 900; Gaps 113;

QY 2 ATCTGGGCTGCTCCAGGAAAAAGCAAAATCTGGAAGTTAAATGTTTG-AGTGAATTTT 60
 Db 2568 ATACGTGTAATGACAAAAATAGGCAAGTTAAAAATTTACAGGCGATGCAATGTTCAA 2627
 QY 61 A-----AATCTT-----GCTGGCGGAGAG-GCCCGCT--CTCCCGC-----GT 97
 Db 2628 ACAGTAAATCAATATCGGGGTGGCGAAGAACTCCAGATGAGATCCCGCGCTGGAGG 2687
 QY 98 ATCAGC-----GTTCTCTCATTTTGAATCGCGGCTCCG-CGGTCTTCGGGCTCAGA 150
 Db 2688 ATCATCCAGCGCGCTCC-CGSAAGAACGATTCGAGAGCCCAACCTTTCATAGAGCGCG- 2745
 QY 151 CCAGCGGAGGAGC-----C-TGTTTGCAATTTTAAGC-----GGGCTGT----- 189
 Db 2746 -----CGTGGATCGAAATCTCGTATGCGAGTTGGCGTGGTGGTTCATTTTC 2800
 QY 190 GAACGCCAGGG-CCGGC-----GGGGCGGGG----- 217
 Db 2801 GAAC-CCCAGAGTCCCGCTCAGAAAGAACTCGTCAAGAGCGGATAGAGCGGATCGGTG 2859
 QY 218 CGAGCGGG-----CCATTTTGAATAAAGAGGCG--TGCC--TTC-CAGGCGAG 261
 Db 2860 CGAATCGGAGCGGCGATACCGTAAAGCACAGGAAAGCGGTCAAGCCCATTCGCCGCCAAG 2919
 QY 262 CTCT-----ATAAGTGACCGCCG-----GGCGAGCGTGGCGCGGTG 299
 Db 2920 CTCTTACGAAATATCAGGATGACCGGATGACCGGATGATGCTGATAGCGTCCGACACCCAG 2979
 QY 300 CAGTCTACTGT-----AGCGGACTTCTT-----TTGGTTTTC----- 331
 Db 2980 CCGGCCACAGTCGATGAATCCAGAAAGGCGCATTTTCCACCATGATATTCCGCAAGCA 3039
 QY 332 --TTTCTCTTTGGGCGACCTCTGCACTCACTCC--CCAGCATGAAGGCGCTGAGCCCGT 387

Db 3040 GGCAATCGCCATGAGTCAAGCAGAGATCTCTCGCGTGGGCGATGGCGCTTGAGCGCTGGC 3099
QY 388 GCG-----CGGCTGCTACAGGCGGTG-TGCT-----GCCTGT-----CGAAGC--CAG 429
Db 3100 GAACAGTTCGGCTGGCGGAGCCCTGATGCTCTTCGTCACAGATCATCTGATCGACAAG 3159
QY 430 TCTGGCCATCGCCCGCGCGCGAG-----GGAAGG 458
Db 3160 ACCGGCTTCATCCGAGTAGTGTCTCGCTCGATCGATGTTTCGCTTGGTGGTGAATGG 3219
QY 459 GC-----CCGG--CAGC-TGAGGAGCGCTG-----AG-CTTGCTGGA----- 492
Db 3220 GCAGGTAGCGGATCAAGCGTATGCGAGCGCGCATGATGATCAGCCATGATGATGATCTTT 3279
QY 493 --CGACATGAAC--ACTG-----CTACTCCGCG-CTGCGGGAATGCTGATC 533
Db 3280 CTGGCAGGAGCAAGGTGAGATGACAGGAGATCTCTGCGCGGCACTTCGCGCCATAGCAG 3339
QY 534 CCGG---AGTCCG---AGAGGCA-----CTCAGCTTAGCCAG-----GT 567
Db 3340 CCAATGCTTCCCGCTTCAGTGACAACTGCGAGCAGAGCTGCGCAAGGAAACGCCCGCTGCT 3399
QY 568 GGAATCC-----TACAGCGCGTCAATCGAC-----TACATTCCTCGAC-CTG--CAGGTA 613
Db 3400 GGCAGCCAGATAGCGCGCTGCTGCTGCTGCAATTCATTCAGGACACCGGACAGGTC 3459
QY 614 GTCTGCGC-----GAGCC---AGCCCT-----CGACCCC----- 641
Db 3460 GGTCTTGACAAAAGAACCGCGCGCCCTGCGCTGACAGCGGAAACAGCGCGGCGATCAGA 3519
QY 642 -----CTGATG-GCCC-----CC-----ACCTTCCCATCCAGACAGCGC- 674
Db 3520 GCAGCGGATGCTGCTGTGTGCGCCAGTATAGCGGAATAGCTCTCCACCGAAGCGCGG 3579
QY 675 -----AGCT-----CGCTCGGAACTGTG----- 694
Db 3580 AGAAGCTGCGTGAATCTCTGTGTTCAATCATGGAACAGATCTCATCTGCTCTCTG 3639
QY 695 -----ATCTCAACAGCAAAAGGAGCTT-----TTGCC-----AC 724
Db 3640 ATCTGATCTTGAATCCCTGCGCCATCAGATCTTGGCGGCAAGAAAGCCATCCAGTTAC 3699
QY 725 T-----GACT---CGGCGGTGTC-----CTGACACTCCAGAGCGAGGT 761
Db 3700 TTTCAGGGCTTCCCAACCTTACAGAGGCGCGCCAGCTGGCAATTCGGTTGCT--T 3757
QY 762 GCTGG-----CGCCCGTCTT-----GCC---TGGGACC-----CGGG 791
Db 3758 GCTGTCATAAACCGCCAGTCTAGCTATGCGCATGTAAGCCCACTGCAAGTACTGTC 3817
QY 792 AACTC-----TCCTG-----CCGGAAGC---CGGACGGCAGG 821
Db 3818 TTCTCTTTGCGTTCGGTTTCCCTTGTCCAGATAGCCAGTAGCTGACATTCATCCGG 3877
QY 822 GATGGCCCAACTTCGCGCTGCCACTTGACTTC-ACCAATCC-CTTCTCT---GGAGA 876
Db 3878 GGTGAGCAGCGTTT-----CTGGGACTGGCTTCTAGTGTTCCGCTTCTCTTATGACAGC 3932
QY 877 CT---AAACCTG-GTCT--CAGGAGCGAAGGAC-----TGTGAACCTGTAGC--- 918
Db 3933 CTTGCGCCCTGAGTCTTGGCGCAGGTAAGCTACATATATGTAATCGG--GCAAT 3990
QY 919 --CTGAAGACCAAGAGTACTGTCGCCACCAAG-----CTGGG-CGACGTC----- 961
Db 3991 CGCTGAATATTC--TTTGTCTCCGACCATCAGGACCTGAGTCGCTGCTCTTTTTCGTGA 4049
QY 962 -----ACCTGCTCCAC----- 974
Db 4050 CATTCAGTTCGCTGCGCTACGCTCTGGAGTGAATGGGGTAATGGGCTACAGCGC 4109
QY 975 CC-----CACCCCAAGTTCT-----AAGGTCTT 998

4110 CCTTTTATGGATTCTGCAAGGAAACTACCCATAATAACAAGAAAGCCCGTCAACGCGCTT 4169
QY 999 TTCAGAGCG-----TGGAGG-----TGTTGG-----AAGGA 1023
Db 4170 CTCAGGGCGTTTATGCGGGGCTGCTATGCTGTTGTTGCTATCTGACTTTTGTCTGACGA 4229
QY 1024 GTGGCTGCTCTC-----CA 1037
Db 4230 GTTCTGCGCTCTGATTTTCCAGTCTGACCACTTTGGGATTATCCCGTGACAGGTCAITCA 4289
QY 1038 AACT-----ATGCC-----AAGGCGGG-----CAGAGCTG---GTCTT--- 1069
Db 4290 GACTGGCTAATGACCCAGTAGGAGCGGATATCATCAACAG-GCTTACCGCTTCTACTG 4348
QY 1070 -----CTGCTCTCC-----TTGG----- 1082
Db 4349 TCGAAGACGTGCGTAACGATATGCTATGCTCTCCCATGCGAGAGTAGGAACTGCCAGGCA 4408
QY 1083 -----AGAAAGGTTCTGTTGC-----CCT---GATTTA----- 1107
Db 4409 TCAATATAAACGAAAGGCTCAGTGAAGACTGGGCTTTCGTTTATCTGTTGTTGTC 4468
QY 1108 --TGAACCTCTAT-----AATAGAGTATAT-----AGTTTTGTACCTTTTTTA--- 1148
Db 4469 GGTGAACGCTCTCTCTGAGTAGGACAAATCCGCGGAGCGGATTTGAACGTTGGAAGCA 4528
QY 1149 -----CAGGAAGTCACTTTCTGTA-----ACAATGCGATGTAT---ATTAAAC- 1189
Db 4529 ACGGCCGAGGGTGGCGGAGGACGCGCCGCGATAACTGCCAGGATCAATTAAGCA 4588
QY 1190 -----TT-----TTTATAAAGTT 1203
Db 4589 GAAGGCCATCTCTGACGAGTGGCTTTTGTGCTTTCTACAACT 4631

RESULT 5
AAU07343
ID AAU07343 standard; protein; 1679 AA.
XX
AC AAU07343;
XX
DT 04-DEC-2001 (first entry)
XX
DE 1-aminocyclopropane carboxylate (ACPC) synthase #12.
XX
KW 1-aminocyclopropane carboxylate synthase; ACPC synthase; brain trauma;
excitatory neurotransmission; neurodegeneration; stroke; nerve damage;
neurodegenerative disease; Alzheimer's disease; depression; epilepsy;
alcohol abuse; cognitive function; memory; learning impairment; human.
XX
OS Homo sapiens.
XX
DN W0200168879-A2.
XX
PD 20-SEP-2001.
XX
PF 14-MAR-2001; 2001WO-BF002857.
XX
PR 14-MAR-2000; 2000US-0189086P.
XX
PS 05-APR-2000; 2000US-0194702P.
XX
PA (PARB) BAYER AG.
XX
PI Ramakrishnan S;
XX
DR WPI; 2001-550286/61.
XX
PT Isolated polynucleotide encoding a human 1-aminocyclopropane-carboxylate
PT (ACPC) synthase, useful for treating brain trauma and neurodegenerative
PT disease (e.g. Alzheimer's disease, depression, epilepsy).
XX
PS Claim 1; Page 231-237; 242pp; English.
XX

CC The invention relates to reagents and methods for regulating excitatory
 CC neurotransmission, and to prevent neurodegeneration. The method involves
 CC the use of an expression vector or a reagent that modulates the activity
 CC of a 1-aminocyclopropane-carboxylate (ACPC) synthase polypeptide. The
 CC reagent is useful for modulating the activity of an ACPC synthase in a
 CC disease such as stroke, a nerve damage or a neurodegenerative disease.
 CC The ACPC synthase polypeptide, polynucleotides and modulators are also
 CC useful for treating brain trauma and neurodegenerative disease (e.g.
 CC Alzheimer's disease, depression, epilepsy). The ACPC synthase modulators
 CC are also useful for treating alcohol abuse and improve cognitive function
 CC and memory of patients with learning impairment. The present sequence
 CC represents the amino acid sequence of human 1-aminocyclopropane-
 CC carboxylate (ACPC) synthase #12, used in the method of the invention
 XX
 XX Sequence 1679 AA;

Query Match 42.6%; Score 3215.5; DB 4; Length 1679;
 Best Local Similarity 42.8%; Pred. No. 1.8e-135; Indels 555; Gaps 102;
 Matches 731; Conservative 0; Mismatches 420;
 QY 2 ATCTGGGTGTCTGCGGAGAGGCCGCTCTCCCC-----GG-----TATCAGC- 103
 Db 1 ATC--GAGTGC-----GGGTA-----TCTGGCTG-----TGGATTGCGCGCGCTCTG 41
 QY 58 TTTAATCTCTGCTGCGGAGAGGCCGCTCTCCCC-----GG-----TATCAGC- 103
 Db 42 CTGAGCGCTGG-AGGC-TGGAACCCCGCCGCCCTTACCCAGGCTTTACTCCACCC 99
 QY 104 --GCTTCTCATTTTGAATCGCGCTCGCGGTCTTCGCG-GTCAGACCCAGCCGAG 160
 Db 100 CGGCTTCGCCCACTGTG---CTGCCCTTCTCGGACCTGGGTGTGCGGAGAGCTGGAG 156
 QY 161 GAAGCTGTGTCGAAT---TTAAGCGGCTG-TGAAGCCCGAGGCC-GGCGGG----- 210
 Db 157 -----ATGTTACCTTCTTAAAGACCTTCAGGGTTCACACCTGTCTGGGCCCA 211
 QY 211 -----CGGCGCGAGGCGG--GCCATTTTGAATA-----AAGAGCGTGCCTTCAG 256
 Db 212 CTTGATGACAGACCTGGGCTAGCCATGGGGAAGATCTGGAAGGAAATGC--TCCAG 269
 QY 257 ---GCAGCTCATTAAGTAC-----CGCGCGCGGAGCGTGC--CGCTTGAGGTAC 307
 Db 270 ARAACTGACAGAGCTGCAGAGCTCCGTGGAGTGGGTGATCTGCCATGATCTCTC 329
 QY 308 TG-----TAGCGGATCTT-----TTTGGT-----TTCT-----TTCT-CTTT 340
 Db 330 TGATACCTCTACCTGTCTCTAGAGGAAGATGATTAATGGTTCTGGGATTCTAGCTGA 389
 QY 341 GGGGCACTCTGGACTCTACCTCCAGCATGAAGCGCTGAG-----CCCGGTGCGC 391
 Db 390 GGAGGGTACAGACTAC-CACATGATGATGATGAGGACAGAACCCAGTG-GC 447
 QY 392 GGTGCTAGAGG-----CGGTGTG-----CTGCTGTGGAAC-GCAGT-CTGGC-- 435
 Db 448 ATCATTAATTGGGACACAGTGAACAACCTGCG---TTTGACCTGTGTCTGGCGG 504
 QY 436 -----CATCGCCCGGCGGAGGAGGGCC-----CGGCA---GCTGA-----G 472
 Db 505 CTGAGTCAGCGCATCATCAGAGGGTGGAGCCATCCCTGTGAGTATGCTGACTGAGG 564
 QY 473 GAGCGCTGAGTTCTCTGGAGCATGAACCACTGCT--ACTCCGCTCGGGA-CTG 529
 Db 565 GGACATCTGTTCTCTCGGAGGAAGTGGCAAGTCTCTGTTCTTACTGCAAGAGCCCA 624
 QY 530 GTACCC-----GGAGTCCCGAGAG-----CACTCAGCT----- 558
 Db 625 GTACCCCTCAGACAGAGAACTGTGTGTCTCTGTAATGGTGTGCTCGCTCTTCTGTGT 684
 QY 559 -TAGCCA-----GGTGAATCTCTACAGCGCGTCTCGACTA---C 595
 Db 685 CTGGCCACAGGTCTGTGTGAGCGGGAGGCTTCTGTGATCCCCACCCCTTACTATGGC 744
 QY 596 ATTCTCGACCTGCAGGT-AGTC---CTGGC-----CGAGCCAGCC-----CCTGGAC 638

RESULT 6
 AA337007
 ID AA337007 standard; protein; 1811 AA.
 XX
 AC AA337007;
 XX
 DT 23-OCT-2003 (revised)
 DT 07-AUG-2003 (first entry)
 XX

Db 745 GCTATACACAGACGCTGTCTCTATGGCAACATCGGCTGCGCTATGTCTACCTGGAC 804
 QY 639 C-----CCCTGATGGGCC-----CCACTTTCCCATCCAGACGCCAGCTCGCTCCGGAA 688
 Db 805 AGTGAGGTGTGAAGGTCAAGGCTCATCTCAT-CAGCC--CCAGAAACCTCTGGGT 861
 QY 689 CTTGTCATCTCCACGA-----CAAAAGGAG--CT-----TTTGGCAC-TGACTC-GGC-- 733
 Db 862 GATGTATCTCTCCCTGAGAGCTACAGAGTACTGTATTTTCCCAAGAGGCAAGGCTG 921
 QY 734 CGTGTCTT-----GACACC-----TCC-----AGAACCGAGTGTGTGCGGC 770
 Db 922 CATGTGATGTGATGAGGTCTACATGCTGTCTCGGTGTTTGAGAACTCTGTTG--GGTACC 979
 QY 771 ----CCTCT--GCTGG-----GACCCC-----GGGACC--TCTCTCCCGAAG 809
 Db 980 CAGTGTCTTAAGCTCTGAAAGGCTCCCTGACCCCGAGAGGCCCAATGATGTGGSCAA 1039
 QY 810 CC-----GAGCGGCGAGG-----ATGGGCCCACTT--CGCCCTGCCCACTTGACTTCA-- 857
 Db 1040 CCAGCAAGGACTTGGGATGCTGGGCTCGGCTTTGGCACGCTGTACAGAAACACAG 1099
 QY 858 -----CCAAATCC-----CTTCTGGAGACTAACTGTGTCTCAGAGCGAGGACTGT 907
 Db 1100 ATGTGGCACTGCGCTGGCTTCCCTCTGCGCTACACGGCCTCAGTGGCTTGGTCCAGT 1159
 QY 908 GAACTGTGAGC-CTGAAGACCCAGAGTACTGTG-----CC-----ACCAGCTGGGC 955
 Db 1160 -ACCAGATGGCACAGCTGTCTCCGGACCGTGAAGTCAACAGGTGTACTCTGCGGAA 1218
 QY 956 GAGCTCACTCTGCT-----CCCAACCC-----TTTTCA-----GAGC----- 977
 Db 1219 AACATGCGCGCTCAAGGTGCCACACCTATGTCTCAGAGAGCTTAGGGATTGGGG 1278
 QY 978 ACCCCCAAGTTCTAAGTCT-----TTTTCA-----GAGC----- 1006
 Db 1279 ATCCCC---TTCTGAGTCTGGGGCTGGCTTCTTCTGCTGCTGACTTGAGAAAGTGA 1335
 QY 1007 -----GTGAGGTGTG-----GAAGGAGTGTCTCTCCAACT-----ATGCC- 1046
 Db 1336 ATGCTGTGTGAGGTGCGGGCTGAGAGGAGTTTCAGGTCTCCCTCCAGAGTACTGCCC 1395
 QY 1047 -AAGGC-----GGCGGAGAGCTGGTC-----TTCTGGTC----- 1075
 Db 1396 AAGGCACTTTGAGGAGGAATGCTGCTCTGGCGCGCTTTTGGACAAACAGGTGCTG 1455
 QY 1076 --TCCTTGGAGAAAGG--TTC--TGTTGC--CCTGATT-----TATGAACTCTA 1116
 Db 1456 CTGTCCTTTG-GCAAGGCTTCCAGTGTAAAGAGCTGGTGGTTTCGCTTTGTCTCTC 1514
 QY 1117 TAATAGATATATAGTTTGTACCT-----TTTTTACAG-- 1151
 Db 1515 AGACCAAGTCCACCGGCTTTG--CCTGGGATGACAGAGGTCCAGCAGGTGCTTGCAGC 1572
 QY 1152 -----GAAG-----GTGACTTTCTGT-----AACAAATGC 1175
 Db 1573 AATCCCAAGTGGCAGAGACCCCGTCCCTCTCAGAGCCAGGACCAAGTACCAACGC 1632
 QY 1176 GATGTATATTAACCTTTTATAAAG 1201
 Db 1633 -AGGTGAGCTGGTCAATGTTCTCGTG 1657

DE Micromonospora carbonacea polyketide synthase (PKS) type I #13.
 XX Macroliide; roseraminin; polyketide; polyketide synthase; PKS; enzyme.
 KW Macroliide; roseraminin; polyketide; polyketide synthase; PKS; enzyme.
 XX Micromonospora carbonacea.
 OS Micromonospora carbonacea.
 XX Key Location/Qualifiers
 FH Misc-difference 452..1811
 FT /note= "Encoded by GCC"
 FT
 XX CA2391131-A1.
 PN 19-NOV-2002.
 XX 26-JUL-2002; 2002CA-02391131.
 XX 26-JUL-2001; 2001US-0307629P.
 XX (SCOP-) ECOPIA BIOSCIENCES INC.
 XX Yang X, Staffa A, Farnet CM;
 PI WPI; 2003-343556/33.
 DR N-PSDB; AAD55823.
 XX Novel isolated polypeptide involved in biosynthesis of macroliides by
 PT microorganisms, useful for biosynthesis of macroliides by microorganisms,
 PT preferably for biosynthesis of roseraminin.
 XX Claim 13; Page 183-189; 206pp; English.
 PS The invention relates to genes and proteins involved in the biosynthesis
 XX of macroliides by microorganisms. In particular it relates to the nucleic
 CC acids forming the biosynthetic locus for roseraminin (a 16-member
 CC macroliide antibiotic) from Micromonospora carbonacea. The invention is
 CC useful for the biosynthesis of macroliides and related chemical structures by
 CC direct manipulation of the proteins involved in the biosynthesis of
 CC roseraminin. It is useful to catalyze certain biochemical reactions, in
 CC vitro or in vivo, to direct or enhance the synthesis or modification of a
 CC polyketide, polyketide substrate or its precursor. The present sequence
 CC is M. carbonacea polyketide synthase (PKS) type I protein. (Updated on 23
 CC -OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1811 RA;
 Query Match 42.1%; Score 3173.5; DB 6; Length 1811;
 Best Local Similarity 46.9%; Pred. No. 1.4e-133;
 Matches 683; Conservative 0; Mismatches 385; Indels 387; Gaps 99;
 QY 2 ATCTGGGGTGTGCTCCAGGAAACCAATCTGGAAGTTAATGTTTGGATGATTTTA 61
 DB 453 ATGCCAGACACCCCGAGCTGAACCGGATCTGACG-----C 490
 QY 62 AATCTTGTGCGGAGAGGCC-----CGC-----CT-----CTCCCGGTATCAGCGCTTCCT 110
 DB 491 GATCTCTGC-----CCAGGAGACGACGCGCGGAGCTGCGGCTCTGCGGTGC-CCTCCT 546
 QY 111 CATTTTGAATCG-----CG--GTC-----CG-----TCTTGGCGGTGAC 151
 DB 547 CCTACCGGG-----CGGTGCGGTGACAAAGGACGAGCGGGGATGTTCTCGGC-----CTTC 601
 QY 152 CAGCGGAGGAG--CCTGTTTGAATTTAAGCGGGTGTGAACCGCCGAG-----GCCG- 204
 DB 602 CGCCAGGAGAGAGACCG--CGCAA-----GTGCTGCACACGAGGAGGTGCGG 651
 QY 205 -GCGGGCGGCGGCGGAGCG--GGCCATTTTGAATAAGAGGCT--GCCTTCAGGGAGG 261
 DB 652 TGCCCCAGCTGGCCCCCGGAGGCGCT-----CGTCGCGGTCTCTGGCCAG 697
 QY 262 CTCATAAG-----TGACCGCGCGG--CGAGC--GTGCGCGGTTGCGAGGT 304
 DB 698 CTCGTCAACTACACAGGTCTGGTCTGCTGTTGTCGCGCGCTGCCACCTT--CGGCT 756

QY 305 CACTGTAGCGGACTTCTTTTGGT---CTTCTCTTTGGGG--ACCT-CTGACTC 357
 DB 757 TCCTGGAGCG--CTACGCGCGGCTCTCCGAGCTGGCCCGGGGACACGACCTCGGTAC-C 813
 QY 358 AC-TCCCCAGCATGAA-----GGCGGTGA-----GCCCGTG-----CGC- 391
 DB 814 ACATCTCGGCTCGGACCTGCGCGCGTGGTGTGTGAGGTGCGGCGCGGTCAACCGCT 873
 QY 392 GGCTGCT-----ACGAG--GCGGTGTGCTGCTGCTG-----GAAACGAGTCT 432
 DB 874 GCGCGCGGCTGACAGGTCTGCGGCTGCGACTGCTCTCTCGGTGGAGTGGAGTCCGCCAC- 932
 QY 433 GGCCA-----TCG-CCCGGGGCGAGGG--AAGGGC-----CCGCGAGC 468
 DB 933 GGCCAGCGGACACCATGCTCGACCCGAAACGCGGATCTGGGCTTCGAGACCACTTC 992
 QY 469 TGAGGAGCGCTGAGCTTGC---TGGAAGACATGAACCA-CTGCTAC---TCCCGCTGC 521
 DB 993 GCGGCGCTCGCGGAGATCGGCTTGTCAAGCGAACCAAGCTGATGCCAAACCGACAC 1052
 QY 522 GGGAACTGGTA-----CCCGGAGTCCCGAGA---GGCA---CTCAGGTTA--GCCAGGTG 568
 DB 1053 CTGACCTGGAGAGGCGCGCGCGGAGTGGTCAATCCACCGCTTACCGCAGCTG 1112
 QY 569 GAAATCTTACAGCGCTC---AT-----CGACTACATCTTC-----GACCTG 607
 DB 1113 G-TCTCCGCAACCGGCGCGGATGAAGACGAGGCGCAACGCTCTCTGTGGGGGGCGAG 1171
 QY 608 CAGGTAGTCTTGGC--CG--AGCCAGCGCTTGAGACCCCTGATGGCCCCCAGCTTCCCATC 664
 DB 1172 C-GGCGTCTCGGCGGCTTCCCA--CCAG--CTGCTGTGCGCGCG--GGSCCAATC 1224
 QY 665 CAGACAGC-----CGAGCTGCTCGCGAACTTGTGATCT--CCAAAGCAAAAGGAGCT 716
 DB 1225 CGTCTGCTGCTCTCCAGCGCGGCAAGGC--GACATCTGCGCTG--GATGGCGCC 1280
 QY 717 TTTGCCACTGAC--TCGGCGGCTC--CTGACACCTCCAGAGCGAGGTGCTGGCGCCGTT 774
 DB 1281 GAGGCC---GTATGACACGCTGCCGAGACTACCGCTT--CTGCT--CGACGAGGCA 1335
 QY 775 CTGCTGGAGACCCCG---GGAACTCTCTCTG---CGAAGCGGAG---GG 817
 DB 1336 C--CAGAAATCCCGGAGTGGAGCGCTTCGGGCGACGCAATTCGGGAGCTGACCGGAG 1393
 QY 818 CAGGATG--GGCCCCCACTTCG---CCCTGCC---CACTTGACTTCAC-----858
 DB 1394 CGAGGACGTGACATGCTTTCGAGACACCCCGCGGGGAGACGTTCCGCGCTCGTCTA 1453
 QY 859 CAAATCCCTTCTCTGAGACT-----AAACCTGTGCTC--AGGAGC-----GAAGGA 903
 DB 1454 CGTGACCCGCAAGAGGACCGCTGCTCACTGCGCTCGACGAGCGGTTTCGAGCACT 1513
 QY 904 CTGTGA-----ACTTGTAG-----CCTGAAGAGCC---AGAGCTAGCTCTG--GCCAC 946
 DB 1514 CTACGACACCGTTACCTGTGATGTCCTGAAGCGCATGCTCGGACGCACTTCGCCAA 1573
 QY 947 CAGCTGGGCGAGT-----CACCC-----TGCTC-----CCACCCCAACCCC 982
 DB 1574 TTACCGGAGGCTGGGAGCAACCGGTTGGTGGTCAAGGGCAAGATCCACCCGAGCT 1633
 QY 983 CAACTTCAAGTCTTTTCAGAGCTGAGGTGTGAAGAGGTGGCTGC---TC--TC- 1035
 DB 1634 GTCGCGCTG-----TACCCGCTGGAGAGGTTCGCCAGGCG--GTCTAGAGCTCCATCA 1687
 QY 1036 CAAACTATGCCAAGC---GGCGGAGAGCTGCTCTTCTGTGCTCTCTTGGAGAAAGTTTC 1092
 DB 1688 CAACTGACACAGGCGAAGGTGCGC--GTGCTCGCTCGCGCGC---GCGAGGGGCTC 1742
 QY 1093 TGTGCGCTGATTTAAGACT--CTATAATAGAGTATATAGGTTTGTACCTTTTACAG 1151
 DB 1743 -GGGGTCCGGAACCGGAGCTGCGGAATGCCATCTTCCGCGATCAACCGCTT---CCG 1798

```

QY 1152 GAAGGTGACTTTCTG 1166
DB 1799 ---GGTGGCGGCTG 1810

RESULT 7
ID ADA15723
XX ADA15723 standard; protein; 2601 AA.
XX AC ADA15723;
XX DT 06-NOV-2003 (first entry)
XX DE C. elegans neuromuscular junction GABA receptor complex subunit #5.
XX KW Nematode;
XX KW neuromuscular junction gamma-aminobutyric acid receptor complex;
XX KW GABA receptor; parasitic plant pathogen; agricultural industry;
XX KW crop protection; soil treatment.
XX OS Caenorhabditis elegans.
XX PN US2003065144-A1.
XX PD 03-APR-2003.
XX PF 24-MAY-2002; 2002US-00156240.
XX PR 09-NOV-1998; 98US-0107272P.
XX PR 08-NOV-1999; 99US-00436063.
XX PA (UTAH ) UNIV UTAH RES FOUND.
XX PI Bamber BA, Jorgensen EM;
XX DR WPI; 2003-540802/51.
XX DR N-PSDB; ADA15724.
XX PT New nematode neuromuscular junction GABA receptor complex, useful for
XX PT crop protection or soil treatment.
XX PS Claim 21; Page 46-52; 84pp; English.
XX CC The present invention relates to a nematode neuromuscular junction gamma-
XX CC aminobutyric acid (GABA) receptor complex. Nematodes are parasitic plant
XX CC pathogens which can cause major damage to crops in the agricultural
XX CC industry. The nematode neuromuscular junction GABA receptor complex is
XX CC useful for crop protection or soil treatment. The present sequence
XX CC represents a Caenorhabditis elegans neuromuscular junction GABA receptor
XX CC complex subunit.
XX SQ Sequence 2601 AA;

Query Match 41.2%; Score 3105; DB 6; Length 2601;
Best Local Similarity 35.9%; Pred. No. 2a-130;
Matches 801; Conservative 0; Mismatches 389; Indels 1040; Gaps 132;

QY 1 GATCTGGGTGC-----TGC--CA-----GGAAAAAGCAAAATCTGGAA 37
DB 364 GATATGACCTCATTAGACTTCTACATGCGTCAACGCTGCAAGACCCCTCGACTAGCC 423
QY 38 GTTATGCTTTGA---GTGATTTT-----AATC-----CTTGCTGGCGGAGG-- 81
DB 424 TTGGGAAGCTTTGATTGGGACTTTTCCAAAGAAATCGACTCACTTACCGTGGAGTAGAC 483
QY 82 -----CCGCGC-----TCTCCCC----- 94
DB 484 TACCTGGATAGACTGTGGAAACCGACAGCTTTTCCCAATGAAGAAATCATTTCTC 543
QY 95 -----GGTATCAGCGC-TTCTCAT-TCTTGAATC----- 123
DB 544 CACTTGGCAACCAACATAACTCGTCTCTGATCGAGGTGATGGAACGGTTTATACT 603

124 ---CGCGGCTCGGGT--CTTCGGGT---CA---GACCAG-----CGGAGGAA 163
604 AGTCAAGATTAAAGTCACTCACTGCAAGCTGTCGAATCGAAGCTGTGTTCCCAATGGA- 662
164 GCCTGTTTGAATTTAAGC--GGGCTGTGAAGCC--CGAGGGCC--GGCGGGGGC----- 212
663 --CTCAACACTGTAAACTGGAATTTGAAAGCTACGGGTACGAGACGAAAGATATCGAC 720
213 -----GGGCGCGAGCGGGCCATTTTGAATAAAGAGGC-----GTGCTTTC 253
721 TACTATTGGGGGAAGAGCGGACTGATTGGAGATTAACGGCTGTCAAGTTTGATACCTTC 780
254 CAG-----GAG-----GCTCTA-TAAGTGA-----CCG---CGCG 281
781 CAGTTGCGGAGTTTTCAGCCACGCTGTATTTTGTGAATACAACTAAAGCCGAGACTCA 840
282 GCGAG-----CGTGGC--GGTTCAGGT-----C 305
841 TCAGAAATACGTACGCTCGGCTGGAAGTAATATTGTTTGAATATGGCCTTCAC 900
306 ACTGT-AGCG--GACTTC-----TTTTGGTTTCTTT----- 334
901 ACTATGAACATCGTCATCCCATCATCTGATCGTCAACATATCTTTGGGTATCATTTGG 960
335 -----CTC-----TTTGGG-----GCAC----- 347
961 TTGAATCGAGAAGCTTCGCGGCTCGAGTTGGATTGGGTGACTACTGTCTCAATG 1020
348 ---CTCTGACTC-ACTCCCCA-----GCATGAAGGCGCT-----GAGCCG 385
1021 ACAACTCTGATCACTACACCAATAATTCGATCCAAAGTGTCTTATGTCGAAGGCTGT 1080
386 G-TGCGGGC-----TGCTACGAGCGGTGTGTGCT-----GTGG--GAAAGC-- 427
1081 GATGTGTTCTTAATTTTGTTCGTAATGTTATTC-GCTGTGTTCTCGAGTACGCCAT 1139
428 -----AGTCTGGCCATCGCCGG--GCCGAGGGAAG--GCC-- 461
1140 AGTATCTTACATGAATAAAGACTGTGCTTCGACGGGAAACGAAAGAAAGCCGCGA 1199
462 ---CGGCAGC---TGAGGAGCGCT-----GAGCTTGTCTGGACGACATGAA--- 501
1200 ACAAAGCAGCGAAACGAGATGCCAATGTTCAACGGGAGCGCGGAGCGGCAATATAA 1259
502 -----CCACT-----GTAC--TC-CGCGCTGGGGA-CTG-GTACCGG 537
1260 TTGATGAAATGACACTTATGTCGAAATTCGACGCTTCGCAAGATATGTACAGC 1319
538 AGTC-----CGAGAGGCAC--TCAGCTTAG--CCA--GGTGAATTC----- 575
1320 TGACTTGTACTTTGCCG---GACACAATTCCTCTATGAATCCATTGATGAGATCCCAGA 1376
576 -----TACAGCGCT---CATCGA--CTACATTCTCGACCTG---CAGGTAGTCC 617
1377 AATTGTGATGCGGCGGATTCGAATGATGCAATCCAGCTCTTGTCAAGACGGGCG 1436
618 -----TGCCG-----AGC-CAGCC-----CCTGAC-----CC 641
1437 ACATACGTATGCGCGCTCCATTCGCGGCGGAAAGGTTCCAGACATGCTGCCA 1496
642 CTGATGCGCCG-----ACC-----TTCCA----- 662
1497 ACGHTGACGCTGCAAAATCGATAGCTTAGCCGATACGTTTCCCATTTCTTCTTC 1556
663 -----TCCAGACAGCC-----GAGCTCGCT---CCGGAATTTGTC-----A 695
1557 TATCTTCAATATAGTCTACTGTTGTATGAAATATCTAAGCTTAACTCGTCGACAA 1616
696 TCTCC-----AAGCAAAAGG--AGCTTTTGCCACTGA-----C-TGGGCC----- 734
1617 GATCCAGAGNACGCAAGTGGCAGAGAT-CCACTGATGCGTATTCGAGCGGCAATC 1675
735 -----GTCTCC-----TGACACC-----TCGAAACGAG-- 759

```


monitoring clinical trials and pharmacogenomic), and in methods of treatment (e.g., therapeutic and prophylactic). NOVX is useful as immunogen to produce antibodies immunospecific for NOVX, as vaccines to screen for potential agonist and antagonist compounds, and as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide is useful in gene therapy, to express NOVX, to detect NOVX mRNA or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The vector is useful for producing non-human transgenic animals. The antibody is useful for isolating, and purifying NOVX and to monitor protein levels in tissue as part of a clinical testing procedure. The present sequence represents a NOVX protein

XX Sequence 1329 AA;

Query Match 41.1%; Score 3103.5; DB 5; Length 1329;
 Best Local Similarity 46.7%; Pred. NO. 1.4e-130;
 Matches 655; Conservative 0; Mismatches 426; Indels 321; Gaps 88;

QY 1 GATCTGGGGTGTGCGCAGAAA-----AGCAAAATCTGGAA----- 37
 DB 23 GCTCT-GCCAGCTACTGGGCAACGACACCCGCATCGCTGGTACCAACCGAGCCCT 81
 QY 38 GTTAATGTTTGTAGTGATTTTAAATCTTGTGGGGAGAGCCCGCTC-TCCCGG 96
 DB 82 GTGGAGGGTGATGAGC-AGGGGGGATCTTC-CTGGCCGAGA-----GCCCTCATCCACG- 133
 QY 97 TATCAGCG-CTTCTCTCATTTTGAATCGC-----GGCTCCCGGTCTTCGGCGT 146
 DB 134 --ACTGCACCTTCATCACCAGGTGGCCCGCAACTCTGGCTGGCATCACACCTAC----- 186
 QY 147 CAGACACCGGAGGAGCCCTTTGCAATTATAGC--GGCTGTGAACCCCGAGCGCG 204
 DB 187 CAGTCTCTGCTGCAATATCCCTTCACTCAGTGCCTCTGGGC-GGGGGTCCCGCGGAC 245
 QY 205 GCGGG-----GGCGGGCCGAGGCGGGCCATTTTGAATGAAGAGCGGTGCTTCCA 255
 DB 246 CCGAGCCTCCCGCGGTGTACCGT-GCGGCGCGCTGGAGCCAGGGGAC--TACTCCCA 302
 QY 256 GCGAGCCTCTATA--AGTGACCGCGGGGAGC-GTGCGC--GCTTGCAGTCTACTGT 310
 DB 303 --CTGTCTCTACCAACGACATACACAGGTGTGTACACCTTCG-TGCTGATGCCAT 359
 QY 311 AGCGGACTCTTTTGGTCTTCTTCTTTGGGGCACT-CTGGAC-----TCACTCCC-- 363
 DB 360 CAATGCTTCAATGCG-----CTGACCC--TGGCTCACAGCTGGCGGTGTACACGCGA 413
 QY 364 -----CAG-CATGAGGCGCTGAGCCCGGT-GCGGGGTGCTACG----- 401
 DB 414 GCGCGTAGCTTTTTCAGACATGATGATGTGTCTATGTGCTCAGATGATCCAGAAAT 473
 QY 402 ---AGCGGTGTGCTGCTG-TCGGAACGACG-----TCTGGCCATCG 440
 DB 474 TTTGGTTATGT-CGACCAGATCAAAGAGCTGTAGGGTGATGGTGACATGGCCAGCA 532
 QY 441 CCC-GGGCCGAGGAAAGGCGCGGAGCTGAGGCGGCTGAGCTTGTGTGACGACATG 499
 DB 533 ACCTGATGTGTGACGAGCAC--CTGCTGT-----GCTGGCCGAGCGGAGGACAAG 585
 QY 500 AACCACTGCTACT-----CC-----CGCC-TGCGGGAACGTGTACCCGAGTCCC 543
 DB 586 GCTGTCAGCGCATGCTGGGTGCTGAGCGCATTTGGGGGGCG---CCCTCAGCCCG 642
 QY 544 ---GAGAGCA--CTCAGCTTA-GCCAG-----GTGGAAT-----CCTACAGCGCTCA 587
 DB 643 CATGCCAGCAATCTCAGTGAATCGAGGAACGTGGCATTTGAGGCTTAC--CTATCA 700
 QY 588 -----TCGACTACAT--TCTGACCTGCA-GGTAGTCTCTG-----GC 621
 DB 701 AGCCGACACAGCTACGTGGGCTGTGACCTGACAGCTTCCAGAGAGGAGGAGGGGTG 760
 QY 622 CGAGCCAGC-CCCTGAGCCCTG-----ATGCGCCCGACCT--TCCCAT---CCAGACAG 671
 DB 761 CCGGCACACCGCCAGGAAGCCCTGCGCAGAACCCCGCCCTGAGCCCGAGCCCGAGCTG 820

QY 672 CCGAGCTCGTCCGGAACTTGTCTCATCTCCAAACGACAAAGAGAGCTTTTGCCACTG--ACT 729
 DB 821 ACCAGC-AGTCCG---CT--TCGGTGCACACCGGGAGGC-----CCATGTTTCT 867
 QY 730 CGGCGGTGTCTGACACTTCAGAAC-GC-AGTGTGGGGCGGCTTCTGCTGGAGACCC 787
 DB 868 CTGTC--GTCTTTCACATCAAGAACAGCGTGGCCCTGCCTCATCCAGCT---GCCCG 922
 QY 788 CGGG---AACTCT--CCTCCCGGAAG-CGCGAGCGCAGGAGTGGGCCCCCAACTTCGCC 841
 DB 923 CGAGTCTATTCTCATCTTCCGGTCTCCCGGTGGCTCCCGCGGTGGCCCGAGACTGCACC 982
 QY 842 TGC--CCACTGTGACTTCAACCAAT-----CCCTTCT-----GGAGCTAATAACCTGGTGC 889
 DB 983 TGCACCTGCTGCTTTC-CGAAATGGCGCTCTTCCACAGCACAGCAACGCGCTCCCGC 1041
 QY 890 TCAGGAGCG--AAGGACTGTGAACCTTGTAGCTTGAAGAG-----CCAG--AGCTAGCTC 939
 DB 1042 CTTGGAGCTGTGGGCGCTG---GCAAGAGCGGTGGCGTGGCCACCCCGCTCATCTT-CGC 1097
 QY 940 TGG--GCACACAGCTG-GGCGACGTCACTCCCTGCTCCACCCCGCCCGAGTTCTAAGCTC 996
 DB 1098 AGGAACAGTGTGTGTGGCTGGGAAACCTG-----ACAGAGTCACTGGCC 1143
 QY 997 TTTTCAGAGCG-----TGAGGTGTGAAGAGTGGCTGCT-----CTCCAAACT 1041
 DB 1144 GTTTGGCTGCGCACTGGGCTGAGCGAGCGAACTGTGGCGCTTGTGGAGCCAGGAG 1203
 QY 1042 ATGCGAAGCGGGGCG--AGAGCTGTCTTC--TGGTCTCTTGGAGAAAGCTTCTGTG 1097
 DB 1204 GGGCCCG--GGAGCTTGGGGCTGGACCTCGGAGGGCTGCCAG--CTCCGCTCCAGCA 1259
 QY 1098 CCCTGATTTATGAACCTCTATAAT-AGAGTATATAGTTTTGT-ACCTTTTACAGGAAG 1155
 DB 1260 GCCCAATGTGAGCGCCCTGCACCTGCCAGCACTTGGGCAATGTGGCGTCT--CATGGAG 1317
 QY 1156 GTGACTTTCTGTAACAATGGA 1177
 DB 1318 -----CTGAGCCT---CGA 1328

RESULT 9
 ABU8255
 ID ABU8255 standard; protein; 1743 AA.
 XX AC ABU8255;
 XX DT 07-JUL-2003 (first entry)
 XX DE Novel human secreted and transmembrane PRO polypeptide #3.
 XX KW Human; secreted and transmembrane protein; PRO; gene therapy;
 XX KW tumor necrosis factor-alpha release; TNF-alpha release;
 XX KW chondrocyte proliferation; chondrocyte differentiation; tumour;
 XX KW adrenal tumour; lung tumour; colon tumour; breast tumour;
 XX KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
 XX CS Homo sapiens.
 XX FN US2003032127-A1.
 XX PD 13-FEB-2003.
 XX PF 26-JUN-2002; 2002US-00183012.
 XX PR 18-SEP-1997; 97US-0059263P.
 XX PR 18-SEP-1997; 97US-0059266P.
 XX PR 17-OCT-1997; 97US-0062250P.
 XX PR 21-OCT-1997; 97US-0063486P.
 XX PR 24-OCT-1997; 97US-0063120P.
 XX PR 24-OCT-1997; 97US-0063121P.
 XX PR 28-OCT-1997; 97US-0063540P.

PR	28-OCT-1997;	97US-0063541P.	PR	11-JUN-1998;	98US-0088863P.
PR	28-OCT-1997;	97US-0063544P.	PR	11-JUN-1998;	98US-0088876P.
PR	28-OCT-1997;	97US-0063564P.	PR	12-JUN-1998;	98US-0089090P.
PR	29-OCT-1997;	97US-0063734P.	PR	12-JUN-1998;	98US-0089105P.
PR	31-OCT-1997;	97US-0063870P.	PR	16-JUN-1998;	98US-0089512P.
PR	31-OCT-1997;	97US-0064103P.	PR	16-JUN-1998;	98US-0089514P.
PR	13-NOV-1997;	97US-0065311P.	PR	17-JUN-1998;	98US-0089538P.
PR	21-NOV-1997;	97US-0066120P.	PR	17-JUN-1998;	98US-0089598P.
PR	24-NOV-1997;	97US-0066466P.	PR	17-JUN-1998;	98US-0089653P.
PR	24-NOV-1997;	97US-0066772P.	PR	17-JUN-1998;	98US-0089908P.
PR	11-DEC-1997;	97US-0069335P.	PR	18-JUN-1998;	98US-0089952P.
PR	12-DEC-1997;	97US-0069425P.	PR	19-JUN-1998;	98US-0090246P.
PR	17-DEC-1997;	97US-0069870P.	PR	22-JUN-1998;	98US-0090252P.
PR	18-DEC-1997;	97US-0068017P.	PR	22-JUN-1998;	98US-0090254P.
PR	10-MAR-1998;	98US-0077450P.	PR	24-JUN-1998;	98US-0090429P.
PR	11-MAR-1998;	98US-0077632P.	PR	24-JUN-1998;	98US-0090435P.
PR	20-MAR-1998;	98US-0077649P.	PR	24-JUN-1998;	98US-0090444P.
PR	20-MAR-1998;	98US-0078986P.	PR	24-JUN-1998;	98US-0090461P.
PR	27-MAR-1998;	98US-0078959P.	PR	24-JUN-1998;	98US-0090535P.
PR	27-MAR-1998;	98US-0079664P.	PR	24-JUN-1998;	98US-0090540P.
PR	31-MAR-1998;	98US-0079786P.	PR	25-JUN-1998;	98US-0090676P.
PR	31-MAR-1998;	98US-0080107P.	PR	25-JUN-1998;	98US-0090678P.
PR	01-APR-1998;	98US-0080327P.	PR	25-JUN-1998;	98US-0090688P.
PR	01-APR-1998;	98US-0080333P.	PR	25-JUN-1998;	98US-0090690P.
PR	08-APR-1998;	98US-0081049P.	PR	25-JUN-1998;	98US-0090694P.
PR	08-APR-1998;	98US-0081070P.	PR	25-JUN-1998;	98US-0090695P.
PR	09-APR-1998;	98US-0081195P.	PR	25-JUN-1998;	98US-0090696P.
PR	15-APR-1998;	98US-0081838P.	PR	25-JUN-1998;	98US-0090699P.
PR	21-APR-1998;	98US-0082568P.	PR	26-JUN-1998;	98US-0090862P.
PR	21-APR-1998;	98US-0082569P.	PR	26-JUN-1998;	98US-0090863P.
PR	22-APR-1998;	98US-0082704P.	PR	26-JUN-1998;	98US-0091010P.
PR	22-APR-1998;	98US-0082797P.	PR	26-JUN-1998;	98US-0091359P.
PR	28-APR-1998;	98US-0083322P.	PR	01-JUL-1998;	98US-0091544P.
PR	29-APR-1998;	98US-0083495P.	PR	01-JUL-1998;	98US-0091478P.
PR	29-APR-1998;	98US-0083496P.	PR	02-JUL-1998;	98US-0091486P.
PR	29-APR-1998;	98US-0083499P.	PR	02-JUL-1998;	98US-0091628P.
PR	29-APR-1998;	98US-0083559P.	PR	02-JUL-1998;	98US-0091628P.
PR	29-APR-1998;	98US-0084366P.	PR	02-JUL-1998;	98US-0091632P.
PR	05-MAY-1998;	98US-0084414P.	PR	02-JUL-1998;	98US-0094006P.
PR	06-MAY-1998;	98US-0084639P.	PR	04-AUG-1998;	98US-0095282P.
PR	07-MAY-1998;	98US-0084640P.	PR	10-AUG-1998;	98US-0095998P.
PR	07-MAY-1998;	98US-0084643P.	PR	10-AUG-1998;	98US-0096012P.
PR	15-MAY-1998;	98US-0085579P.	PR	17-AUG-1998;	98US-0096757P.
PR	15-MAY-1998;	98US-0085580P.	PR	17-AUG-1998;	98US-0096867P.
PR	15-MAY-1998;	98US-0085582P.	PR	17-AUG-1998;	98US-0096891P.
PR	15-MAY-1998;	98US-0085700P.	PR	17-AUG-1998;	98US-0096897P.
PR	18-MAY-1998;	98US-0086023P.	PR	18-AUG-1998;	98US-0096949P.
PR	22-MAY-1998;	98US-0086392P.	PR	18-AUG-1998;	98US-0096959P.
PR	22-MAY-1998;	98US-0086486P.	PR	18-AUG-1998;	98US-0097022P.
PR	28-MAY-1998;	98US-0087098P.	PR	26-AUG-1998;	98US-0097952P.
PR	28-MAY-1998;	98US-0087208P.	PR	26-AUG-1998;	98US-0097955P.
PR	02-JUN-1998;	98US-0087609P.	PR	26-AUG-1998;	98US-0097971P.
PR	02-JUN-1998;	98US-0087759P.	PR	26-AUG-1998;	98US-0097974P.
PR	03-JUN-1998;	98US-0087827P.	PR	26-AUG-1998;	98US-0098014P.
PR	04-JUN-1998;	98US-0088025P.	PR	26-AUG-1998;	98US-0098116P.
PR	04-JUN-1998;	98US-0088028P.	PR	01-SEP-1998;	98US-0098723P.
PR	04-JUN-1998;	98US-0088029P.	PR	01-SEP-1998;	98US-0098803P.
PR	04-JUN-1998;	98US-0088033P.	PR	02-SEP-1998;	98US-0098821P.
PR	04-JUN-1998;	98US-0088033P.	PR	02-SEP-1998;	98US-0098843P.
PR	04-JUN-1998;	98US-0088328P.	PR	02-SEP-1998;	98US-0099602P.
PR	05-JUN-1998;	98US-0088167P.	PR	03-SEP-1998;	98US-0099741P.
PR	05-JUN-1998;	98US-0088202P.	PR	10-SEP-1998;	98US-0099754P.
PR	05-JUN-1998;	98US-0088211P.	PR	10-SEP-1998;	98US-0099763P.
PR	05-JUN-1998;	98US-0088217P.	PR	10-SEP-1998;	98US-0099763P.
PR	09-JUN-1998;	98US-0088655P.	PR	10-SEP-1998;	98US-0099912P.
PR	09-JUN-1998;	98US-0088722P.	PR	10-SEP-1998;	98US-0099912P.
PR	10-JUN-1998;	98US-0088738P.	PR	10-SEP-1998;	98US-0100388P.
PR	10-JUN-1998;	98US-0088740P.	PR	15-SEP-1998;	98US-0100662P.
PR	10-JUN-1998;	98US-0088811P.	PR	15-SEP-1998;	98US-0100664P.
PR	10-JUN-1998;	98US-0088824P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;	98US-0088825P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;	98US-0088826P.	PR	17-SEP-1998;	98US-0100683P.
PR	11-JUN-1998;	98US-0088866P.	PR	17-SEP-1998;	98US-0100684P.

PR	17-SEP-1998;	98US-0100919P.	Query Match	41.1%;	Score 3101.5;	DB 6;	Length 1743;
PR	17-SEP-1998;	98US-0100930P.	Best Local Similarity	41.8%;	Pred. No. 2.1e-130;		
PR	18-SEP-1998;	98US-0100849P.	Matches	725;	Conservative	0;	Mismatches 421; Indels 587; Gaps 102;
PR	18-SEP-1998;	98US-0101014P.					
PR	18-SEP-1998;	98US-0101068P.					
PR	23-SEP-1998;	98US-0101471P.					
PR	23-SEP-1998;	98US-0101472P.					
PR	23-SEP-1998;	98US-0101475P.					
PR	23-SEP-1998;	98US-0101477P.					
PR	24-SEP-1998;	98US-0101738P.					
PR	24-SEP-1998;	98US-0101739P.					
PR	24-SEP-1998;	98US-0101743P.					
PR	24-SEP-1998;	98US-0101922P.					
PR	25-SEP-1998;	98US-0101786P.					
PR	29-SEP-1998;	98US-0102407P.					
PR	29-SEP-1998;	98US-0102409P.					
PR	29-SEP-1998;	98US-0102330P.					
PR	29-SEP-1998;	98US-0102331P.					
PR	30-SEP-1998;	98US-0102487P.					
PR	30-SEP-1998;	98US-0102570P.					
PR	30-SEP-1998;	98US-0102571P.					
PR	01-OCT-1998;	98US-0102684P.					
PR	01-OCT-1998;	98US-0102687P.					
PR	02-OCT-1998;	98US-0102965P.					
PR	06-OCT-1998;	98US-0103258P.					
QY	1	GATCTGGG-----GTGCTGCCAGGAAAGCAAAATTTCTGGAAGTTAATGGTTTGTAG 52					
DB	12	GGACAGGGCAGGAGAGAGAACTGCAAGA-CATATTTGTTCCAAAATGGCATC--- 67					
QY	53	TGATTTTITA-----AATCTTTCTGGCGGAGGCGCCCTCT---CCCCGGTAT-CAG 102					
DB	68	TTACCTTTATGAGTACTCTTTGCTGTG-----GCCCTGTGCTCCAACTACTATG 118					
QY	103	CGCTTCTCTCATTTTGAATCC---GGGGTCCGGCTCTTTGGCGTC---AGACACAGCG 157					
DB	119	TGTGTCGCCGGCAATGCCCCAGTGCATACCCCGCCCTTCTCTCACAAGAGCACCC- 177					
QY	158	GAGGAAGCC-----TGTTTGCAAATTAAGCGGGCTGTG---AAGCCCGAGG---GCCGG 205					
DB	178	-----CTGCCCTCACAGGTGTAATCCCTCAACACCGACTTTGCCCTCCGCTATACCCGAG 233					
QY	206	CGGG-----GGCGGGCGGCGGGCCATTT-----TGAATAAAGAGGCGTGC 249					
DB	234	CTGGTTTGGAGACCCCGAGTCAGAACATCTTCTCTCCCTGTGAGTGTCTCCACTTC 293					
QY	250	CT-TCCAGG-----AGGCTCTATAAGTGACC-----GCC--- 278					
DB	294	CTGGCCATGTCTCTCCCTTGGGGCCCACTCACTCAACAGACCCAGATTCTCCAGGGCCTG 353					
QY	279	-GGCGCGAGGTGGCGGC-----GT-TGCAGGTCACTGTAGCGGACTTCTTTTGGTT 328					
DB	354	GGCTTCAACCTCACACACACAGAGTCTGCCATCCACGAGGCTTCCAGCACCTGGTT 413					
QY	329	TTCTTTCTCTTTGG-----GGC-----ACCT---CTGGAC-----TCATCTCCCGAG 367					
DB	414	CACCTCACTGACTGTTCCTCAGCAAGACCTGACCTTGAAGATGGGAAAGTGCCCTCTTCGTG 473					
QY	368	ATGAAGCGCTGACCCGCTGC---GC-----GGC---TGCTACGAGCGGCTGTGCT 413					
DB	474	AAGAAGAGCTG---CAGCTCGAGGCAAAATTTCTTGGGCAATGTCAAGAGCTGTATGAA 530					
QY	414	GC-----CTGT-----CGGA-----ACGAGTGTGGCCATCGCCGGGGCGGAGG---AAG 457					
DB	531	GCAAGAGTCTTTTCTACAGATTTCTC-CAACCCCTCCATTTGCCAGG---CGAGGATCAAC 587					
QY	458	GGCCGCGAGCTGAGGCGCTGAGTGTGTGACGACATGAACACTGC-----TACTC 513					
DB	588	AGCCATGTGAANAAGAACCCAGGAGGTTGTAGACATATCCAGGSCCTTGACCTT 647					

QY	514	CCGCTGC---GGAACTGGT---ACCGGAGTC-----CC-----GAGAGGCAC--T 553					
DB	648	CTGACGCCCATGGTTCTGGTGAATCAATTTCTTTAAAGCCAAAGTGGGAGAGCCCTTT 707					
QY	554	CAGCTT-----AG-----CAAGTGG-----AA 571					
DB	708	CACCTTGAATATACAGAAAGAACTTCCCATTTCTGGTGGCGAGCAGGTCACTGTGCAA 767					
QY	572	ATCC-----TACAG---CGCGT-----CATCG-----ACTAC 595					
DB	768	GTCCCATGATGCACCAAGAGAGAGTTCCTTTTGGGGTGGGATAAGAGTGAATGC 827					
QY	596	ATTCTCGACCTGCAGGTAG---TCCTGGCGGA---AGCC---CCTGGACC---CCCT--- 643					
DB	828	TTT---GTGCTGCGAGATGAGATTAAGGGAGATGCCGTGCCCTCTTTTGTCTCCCTAGC 884					
QY	644	-----GAT---GGCCCC---CA---CCTTCCATCCAGACAGCGCA----- 675					
DB	885	AAGGCAAGATGAGGCAACTGGAAACAGGCTTGTGACCCAGAACACTGATAAAGTGGAGC 944					
QY	676	-GCTCGCTCCGGAACCTG-----TCATCTCCCAACGACAAAGAGAGCTTTTGC 721					
DB	945	CACCTCACTCCAGAAAGGTGATAGAGTGTTCATCCCC-----AGATTTT-C 991					
QY	722	CA---CTGACTCGGCCCTGTCTGACACCTCCAGAACGCGAGTCTGTGCGCCCTCTGTC 778					
DB	992	CATTTCTGCTCTCTAC-AACTGGAACCACTCTCCCGAAGATG---GGCATCCAAAATGC 1048					
QY	779	CTGGGACC-----CCGG-----GAACCTCTCTCCCGG----- 806					
DB	1049	CTTTGCAAAAATGCTGATTTTCTGGAAATGCAAGAGAGACTC-CCTGAGGTTCTA 1107					
QY	807	-----AAGCGGACGGCAG---GGATG-----GGCC----- 829					
DB	1108	AAGCAACCCCAAGCTGTGCTGGATGTCAGTGAAGAGGCGCACTGAGGCCACAGCAGTA 1167					
QY	830	-----CCAACTTC---GGCC-----TG---CCACTTGNCTTCAACCAATCCCTT 868					
DB	1168	CCACCAACCAATTCATAGTCGATCGATCGAAGATGTCCTCTT-ACCTCACTGTCTCTTC 1226					
QY	869	CCTGAGACTAACTCTGTGCT-----CA---GGAGCAAGAGTGTGAAT----- 912					
DB	1227	AATAGGACCT--TCCTGATGATGATTACAAATAAGCCACAGAGGTATTTCTCTTTCTAG 1284					
QY	913	-----TGTAAGAGCCAGAGTGTGCTGCCC-----ACCAGCTGGGCGAC 958					
DB	1285	GGAAAGTGGAAATCCCACTAAATCTTAGTGGGAAATGGCCTGTTAACTGATG----- 1339					
QY	959	GTCACCTGTCT-----CCACCCCAACCC-----CAAGTTCTAAGG--TC 996					
DB	1340	---CACTTGTCTAATGCAACAGAAATAACCAACATCCCTCTTCTTCTGTTCTGAGGTCG 1397					
QY	997	TTTTCAAGC-GTGGAGGTG-----TGAAGAGGTGGCT-----GC-----TCTCC 1036					
DB	1398	ATTTGACCCCACTGGAGCTGGATTCGCTGGCAGGATGCCACTTCCAAAGGCTCAATCACC 1457					
QY	1037	AAACTAT-----GCCAAGCC-----GGCAGAG--CTG 1064					
DB	1458	AAACATCAACAGGAGCCCGAGTCAAGCCACACCCATTAACCCAGTCAGTGCCTT 1517					
QY	1065	GTCTTCTGG-TCTCTTGG-AGAAAGGTTC-----TGTTC-----CCTGATT--A 1107					
DB	1518	TTCCACAAATTTCTCCAGGTAACTAGTTTCAATGGATGTTGCTGGGTTACCATATTCCA 1577					
QY	1108	T-----GAACCTATATAGAG-----TATATAGGTTTGTACCTTTTTCAGG 1152					
DB	1578	TTCTTGGGGTCCCGAATGGAATAGCCAAACCCAGGTTAGGACCTCTATTGAGA 1637					
QY	1153	-----AAGGTGACTTTCTGTAAACAAATCGATGTATATTAATTTTATAAAA 1200					
DB	1638	ATTACATAACACACTTCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1690					

RESULT 10
ABU90134
ID ABU90134 standard; protein; 1743 AA.
XX
AC ABU90134;
XX
DT 11-AUG-2003 (first entry)
XX
DE Novol human secreted and transmembrane PRO protein #3.
XX
KW Human; gene therapy; tissue typing; tumour; chondrocyte proliferation;
KW Chondrocyte differentiation; tumour necrosis factor-alpha release;
KW affinity purification.
XX
OS Homo sapiens.
XX
PN US2003036147-A1.
XX
PD 20-FEB-2003.
XX
PF 02-JUL-2002; 2002US-00187741.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066456P.
PR 11-DEC-1997; 97US-0066772P.
PR 12-DEC-1997; 97US-0068335P.
PR 17-DEC-1997; 97US-0069425P.
PR 18-DEC-1997; 97US-0069870P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 05-MAY-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087038P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.

```

PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099784P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101781P.
PR 16-SEP-1998; 98US-0101933P.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 23-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102467P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103288P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-0016897P.

Query Match 41.1%; Score 3101.8; DB 6; Length 1743;
Best Local Similarity 41.8%; Pred. No. 2.1e-130;
Matches 725; Conservative 0; Mismatches 421; Indels 587; Gaps 102;

Qy 1 GATCTGG-----GTCTGCCAGGAAAGCAATTCGGAAGTTAATGTTTGGAG 52
Db 12 GGCACGGCAGAGAGAGAACCTTGCMAAGH-CATATTTGTTCCAAATGGCATC--- 67
Qy 53 TGATTTTA-----AATCCTTGTGGCGGAGAGCCCGCTCT---CCCCGGTAT-CAG 102
Db 68 TTACCTTTATGGAGTACTCTTTGCTGTG-----GCCCTGTGTCTCAATCTACTG 118
Qy 103 CGCTTCCTCATTTTGAATCC---GGGCTCCGGGCTCTTCGGGCTC---AGACCCGCG 157
Db 119 TGTGTCCCGGCAATGCCCCAGATGATACCCCGCCCTTCTTCCAAAGAGACCC- 177

158 GAGGAAGCC-----TGTTTGAATTAAGCGGCTGTG---AACGCCCAGG--GCCGG 205
178 ----CTGCTCACAGGTGTTTCCCTCAACACGACTTTCGCTTCCGCTTATACCGAGG 233
206 CGGG-----GGCGGGGCGAGCGGGGCCATTT-----TGAATAAAGAGGCGTGC 249
234 CTGGTTTTTGGAGAGCCCGAGTCAGAACATCTTCTCTCTCCCTGTGAGTGTCTCCACTTC 293
250 CT-TCCAGGC-----AGGCTCTATTAAGTGACC-----GCC-- 278
294 CTGGCCATGCTCTCCCTTGGGCCCCACTCAGTCACCAAGACCAGCATCTCCAGGGCGTG 353
279 -GGGCGAGCGTGGCGC-----GT-TGAGGTCACTGTAGCGGACTTTCTTTGGTT 328
384 GGCCTTCAACCTTCACACACACACACAGAGTCTGCCATCCACAGGGCTTCCAGACCTGGTT 413
329 TTCTTTCTCTTTGG-----GGC---ACCT---CTGGAC-----TCACCTCCCCAGC 367
414 CACTCACTGACTGTTCCAGCAAGACCTTGACCTTGAAGATGGGAAGTGCCTTCTGTC 473
368 ATGAASGCGCTGAGCCCGGTGC--GC-----GGC--TGCTAGAGCGGCTGTGCT 413
474 AAGAAGGAGCTG---CAGCTGAGGCAAAATTTCTGGGCAATGTCAAGAGGCTGTATGAA 530
414 GC-----CTGT---CGGA-----ACGCACTCTGGCCATCGCCGGGCGCAGGG--AAG 457
531 GCAGAAAGTCTTTTCTACAGATTTCTC-CAACCCCTCATTGCCAGG--CGAGGATCAAC 587
458 GGCCCGGCGAGCTGAGGAGCGCTGAGCTTGTGTGACGACATGAACACTGC-----TACTC 513
588 AGCCATGTGAAAAGAGAGACCCAGGGAAGTTGTAGACATAATCCAAAGGCTTGAACCTT 647
514 CGGCTGC--GGGAACCTGTT---ACCGGAGTC-----CC-----GAGAGGAC--T 553
648 CTGACGGCCATGTTCTGTTGAATCAATTTCTTTAAAGCAAGTGGGAGAGACCTTT 707
584 CAGCTT-----AG-----CGAGTGG-----AA 571
708 CACCTTGAATATACAAAGAAAGTCTCCCATTTCTGTGGGCGAGAGAGTCACTGTGCAA 767
572 ATCC-----TACAG--CGCCT-----CATCG-----ACTAC 598
768 GTCCCATGATGCACCAAGAGAGTGTGCTTTTGGGTGATACAGAGTGAAGTGC 827
596 ATTCTCGACCTGCAGGTAG--TCCTGGCGA-GCC--AGCC--CCTGGACC--CCCT--- 643
828 TTT---GTGCTGCAGATGGATTACAGGAGATGCCGTGGCTCTTTTCTCTCCCTAGC 884
644 -----GAT--GGCCCC-----CA--CCTTCCCATCCAGACAGCCGA----- 675
888 AAGGCAAGATGAGGCAACTGGAACAGGCTTGTGAGCCAGAACACTGATAAAGTGGAGC 944
676 -GCTCGCTCCGGAACCTTG-----TCATCTCCAACGACAAAAGAGCTTTTTCG 721
945 CACTCACTCCAGAAAGGTGATAGAGTGTTCATCCCC-----AGATTTT-C 991
722 CA---CTGACTCGCCGTGCTCTGACACCTCCAGAACGAGGTGCTGGGCGCGCTTCTGC 778
992 CATTTCTGCTCTCTAC-AACTGGAACCATCTCCCGAAGATG--GGCATCCAAAATGC 1048
779 CTGGACC-----CCGG-----GAACCTCTCTCTGCGG----- 806
1049 CTTTGACAAATCTGATTTTCTGGAATTCGAAGAGAGAGATC-CTTGCAGGTTCTTA 1107
807 ----AAGCCGAGCGCAG-----GGATG-----GGCC----- 829
1108 AAGCAACCCACAGGCTGTGCTGATGTCAAGAGAGGCACTGAGGCGCACAGCACTA 1167
830 -----CCAACTTC--GCC-----TG--CCCACTTCACTTCACTCAATCCCTT 868
1168 CCAACCCAGTTCATAGTGGATCGATCGAAGGATGGTCCCTCTT-ACCTTCACTGTCTCTTC 1226

```

QY 869 CTGGAGCTAAACCTGGTCT-----CA---GGAGCGAAGGACTGTGAAC----- 912
Db 1227 AATAGGACCT--TCTGATGATGATTACAATAAAGCCACAGACGGTATTCTCTTCTAG 1284
QY 913 ----TGTAAG-----CCTGAAGAGCCAGAGCTAGCTCTGGCC----ACCAGCTGGCGGAC 958
Db 1285 GGAAGTGGAAAAATCCCACTAAATCTTAGTGGGAATGCGCTGTTAACTGATGG----- 1339
QY 959 GTCAACCTGTCT-----CCACCCACACCCC-----CAAGTCTTAAGG--TC 996
Db 1340 --CACATTGCTAATSCACAAGAAATAACAACCAACATCCCTCTTCTGTTCTGAGGTGC 1397
QY 997 TTTTCAGAGC-GTGAGGTG-----TGGAAGGAGTGGCT-----GC-----TCTCC 1036
Db 1398 ATTTGACCCAGTGGAGTGGATTGCGTGGCAGGATGCCACATCCCAAGGCTCAATCAAC 1457
QY 1037 AAATAT-----GCAAGGC-----GGC-----GGCAGAG--CTG 1064
Db 1458 AAACCATCACAGGAGCCCAAGTCCAGCAAGCCCAACACCCCAATTAACCCAGTCACTGCGCCT 1517
QY 1065 GTCTTCTGG-TCTCTTGG-AGAAAGTTC-----TGTTGC-----CCTGATTT--A 1107
Db 1518 TTCCAAATTTCTCCAGGTAACTAGCTTCATGGGATGTTGCTGGTTACCATATTTCCA 1577
QY 1108 T-----GAACTCTATAATAGAG-----TATATAGTGTGTTGACCTTTTACAGG 1152
Db 1578 TTCTTTGGGCTCCAGGAATGGAATAGCAACCCAGGTTAGGCACCTCTATTGCAGA 1637
QY 1153 -----AAGTCACTTCTGTAACTGCGATGATATTAATTAATTTTATATAA 1200
Db 1638 ATTACATTAACATCTCAATAAATAAATATGATTCATAAAAAAAAAAAAA 1690

RESULT 11
ABU96436
ID ABU96436 standard; protein; 1743 AA.
AC ABU96436;
DT 25-JUL-2003 (first entry)
DE Novel human secreted and transmembrane protein #3.
XX Human; secreted and transmembrane protein; PRO; transgenic animal;
KW knockout; chromosome identification; tissue typing; tumour;
KW chondrocyte proliferation; chondrocyte differentiation;
KW tumor necrosis factor-alpha release stimulator.
XX Homo sapiens.
XX US2003036144-A1.
PD 20-FEB-2003.
XX 01-JUL-2002; 2002US-00187601.
XX 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059263P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.

PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0089870P.
PR 18-DEC-1997; 97US-0089871P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077866P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 01-APR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 08-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 09-APR-1998; 98US-0081070P.
PR 15-APR-1998; 98US-0081195P.
PR 21-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 22-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 28-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 18-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.

PR	19-JUN-1998;	98US-0089952P;
PR	22-JUN-1998;	98US-0090246P;
PR	22-JUN-1998;	98US-0090252P;
PR	22-JUN-1998;	98US-0090254P;
PR	22-JUN-1998;	98US-0090402P;
PR	24-JUN-1998;	98US-0090435P;
PR	24-JUN-1998;	98US-0090444P;
PR	24-JUN-1998;	98US-0090461P;
PR	24-JUN-1998;	98US-0090534P;
PR	24-JUN-1998;	98US-0090540P;
PR	25-JUN-1998;	98US-0090676P;
PR	25-JUN-1998;	98US-0090678P;
PR	25-JUN-1998;	98US-0090688P;
PR	25-JUN-1998;	98US-0090690P;
PR	25-JUN-1998;	98US-0090694P;
PR	25-JUN-1998;	98US-0090695P;
PR	25-JUN-1998;	98US-0090696P;
PR	26-JUN-1998;	98US-00105413;
PR	26-JUN-1998;	98US-0090862P;
PR	26-JUN-1998;	98US-0090863P;
PR	26-JUN-1998;	98US-0091101P;
PR	01-JUL-1998;	98US-0091359P;
PR	01-JUL-1998;	98US-0091544P;
PR	02-JUL-1998;	98US-0091478P;
PR	02-JUL-1998;	98US-0091486P;
PR	02-JUL-1998;	98US-0091626P;
PR	02-JUL-1998;	98US-0091628P;
PR	02-JUL-1998;	98US-0091632P;
PR	02-JUL-1998;	98US-0094000P;
PR	24-AUG-1998;	98US-0094008P;
PR	10-AUG-1998;	98US-0095282P;
PR	10-AUG-1998;	98US-0095989P;
PR	10-AUG-1998;	98US-0096012P;
PR	17-AUG-1998;	98US-0096757P;
PR	17-AUG-1998;	98US-0096766P;
PR	17-AUG-1998;	98US-0096867P;
PR	17-AUG-1998;	98US-0096891P;
PR	17-AUG-1998;	98US-0096894P;
PR	18-AUG-1998;	98US-0096949P;
PR	18-AUG-1998;	98US-0096959P;
PR	18-AUG-1998;	98US-0097022P;
PR	26-AUG-1998;	98US-0097552P;
PR	26-AUG-1998;	98US-0097954P;
PR	26-AUG-1998;	98US-0097955P;
PR	26-AUG-1998;	98US-0097971P;
PR	26-AUG-1998;	98US-0097974P;
PR	26-AUG-1998;	98US-0098014P;
PR	01-SEP-1998;	98US-0098716P;
PR	01-SEP-1998;	98US-0098723P;
PR	02-SEP-1998;	98US-0098803P;
PR	02-SEP-1998;	98US-0098821P;
PR	02-SEP-1998;	98US-0098842P;
PR	03-SEP-1998;	98US-00989602P;
PR	10-SEP-1998;	98US-0099741P;
PR	10-SEP-1998;	98US-0099754P;
PR	10-SEP-1998;	98US-0099763P;
PR	10-SEP-1998;	98US-0100683P;
PR	10-SEP-1998;	98US-0098112P;
PR	15-SEP-1998;	98US-0100388P;
PR	16-SEP-1998;	98US-0100662P;
PR	16-SEP-1998;	98US-0100664P;
PR	16-SEP-1998;	98US-0101751P;
PR	16-SEP-1998;	98WO-10319330;
PR	17-SEP-1998;	98US-0100683P;
PR	17-SEP-1998;	98US-0100684P;
PR	17-SEP-1998;	98US-0100919P;
PR	18-SEP-1998;	98US-0100930P;
PR	18-SEP-1998;	98US-0100849P;
PR	18-SEP-1998;	98US-0101014P;
PR	18-SEP-1998;	98US-0101068P;
PR	23-SEP-1998;	98US-0101471P;
PR	23-SEP-1998;	98US-0103472P;
PR	23-SEP-1998;	98US-0103475P;
PR	23-SEP-1998;	98US-0103477P;
PR	24-SEP-1998;	98US-0103738P;

Novel human secreted and transmembrane protein #3.

Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy; chondrocyte stimulator; tumour; adrenal tumour; lung tumour; colon tumour; breast tumour; prostate tumour; rectal tumour; cervical tumour; liver tumour; TNF-alpha release; tumour necrosis factor alpha release; chondrocyte cell proliferation; chondrocyte cell differentiation; pharmaceutical; diagnostic; biosensor; bioactor.

Homo sapiens.

US200301353-A1.

16-JAN-2003.

19-JUN-2002; 2002US-00175737.

18-SEP-1997; 97US-0059263P.
18-SEP-1997; 97US-0059266P.
17-OCT-1997; 97US-0062250P.
21-OCT-1997; 97US-0063486P.
24-OCT-1997; 97US-0063120P.
24-OCT-1997; 97US-0063121P.
28-OCT-1997; 97US-0063540P.
28-OCT-1997; 97US-0063541P.
28-OCT-1997; 97US-0063544P.
28-OCT-1997; 97US-0063564P.
29-OCT-1997; 97US-0063734P.
31-OCT-1997; 97US-0063870P.
31-OCT-1997; 97US-0064103P.
13-NOV-1997; 97US-0065311P.
21-NOV-1997; 97US-0066120P.
24-NOV-1997; 97US-0066466P.
24-NOV-1997; 97US-0066772P.
11-DEC-1997; 97US-0069335P.
12-DEC-1997; 97US-0069425P.
17-DEC-1997; 97US-0069870P.
18-DEC-1997; 97US-0068017P.
10-MAR-1998; 98US-0077450P.
11-MAR-1998; 98US-0077632P.
11-MAR-1998; 98US-0077649P.
20-MAR-1998; 98US-0078886P.
20-MAR-1998; 98US-0078939P.
27-MAR-1998; 98US-0079664P.
27-MAR-1998; 98US-0079786P.
31-MAR-1998; 98US-0080107P.
31-MAR-1998; 98US-0080194P.
01-APR-1998; 98US-0080327P.
01-APR-1998; 98US-0080333P.
08-APR-1998; 98US-0081049P.
08-APR-1998; 98US-0081070P.
09-APR-1998; 98US-0081195P.
15-APR-1998; 98US-0081838P.
21-APR-1998; 98US-0082568P.
21-APR-1998; 98US-0082569P.
22-APR-1998; 98US-0082704P.
22-APR-1998; 98US-0082797P.
28-APR-1998; 98US-0083322P.
29-APR-1998; 98US-0083495P.
29-APR-1998; 98US-0083496P.
29-APR-1998; 98US-0083499P.
29-APR-1998; 98US-0083559P.
05-MAY-1998; 98US-0084366P.
06-MAY-1998; 98US-0084414P.
07-MAY-1998; 98US-0084639P.
07-MAY-1998; 98US-0084640P.
07-MAY-1998; 98US-0084643P.
15-MAY-1998; 98US-0085579P.
15-MAY-1998; 98US-0085580P.
15-MAY-1998; 98US-0085582P.
15-MAY-1998; 98US-0085700P.
18-MAY-1998; 98US-0086023P.

PR	22-MAY-1998;	98US-0086392P.	PR	18-AUG-1998;	98US-0096959P.
PR	22-MAY-1998;	98US-0086486P.	PR	18-AUG-1998;	98US-0097022P.
PR	28-MAY-1998;	98US-0087098P.	PR	26-AUG-1998;	98US-0097952P.
PR	28-MAY-1998;	98US-0087208P.	PR	26-AUG-1998;	98US-0097954P.
PR	02-JUN-1998;	98US-0087609P.	PR	26-AUG-1998;	98US-0097955P.
PR	02-JUN-1998;	98US-0087759P.	PR	26-AUG-1998;	98US-0097971P.
PR	02-JUN-1998;	98US-0087827P.	PR	26-AUG-1998;	98US-0097974P.
PR	04-JUN-1998;	98US-0088025P.	PR	01-SEP-1998;	98US-0098014P.
PR	04-JUN-1998;	98US-0088029P.	PR	01-SEP-1998;	98US-0098716P.
PR	04-JUN-1998;	98US-0088033P.	PR	02-SEP-1998;	98US-0098723P.
PR	04-JUN-1998;	98US-0088326P.	PR	02-SEP-1998;	98US-0098803P.
PR	05-JUN-1998;	98US-0088467P.	PR	02-SEP-1998;	98US-0098821P.
PR	05-JUN-1998;	98US-0088820P.	PR	09-SEP-1998;	98US-0098843P.
PR	05-JUN-1998;	98US-0088821P.	PR	10-SEP-1998;	98US-0099602P.
PR	05-JUN-1998;	98US-0088821P.	PR	10-SEP-1998;	98US-0099741P.
PR	09-JUN-1998;	98US-0088722P.	PR	10-SEP-1998;	98US-0099754P.
PR	10-JUN-1998;	98US-0088738P.	PR	10-SEP-1998;	98US-0099763P.
PR	10-JUN-1998;	98US-0088740P.	PR	15-SEP-1998;	98US-0100388P.
PR	10-JUN-1998;	98US-0088811P.	PR	16-SEP-1998;	98US-0100662P.
PR	10-JUN-1998;	98US-0088824P.	PR	16-SEP-1998;	98US-0100664P.
PR	10-JUN-1998;	98US-0088825P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;	98US-0088826P.	PR	16-SEP-1998;	98US-0101751P.
PR	11-JUN-1998;	98US-0088861P.	PR	16-SEP-1998;	98US-0101930P.
PR	11-JUN-1998;	98US-0088863P.	PR	17-SEP-1998;	98US-0100919P.
PR	11-JUN-1998;	98US-0088863P.	PR	17-SEP-1998;	98US-0100930P.
PR	11-JUN-1998;	98US-0088909P.	PR	18-SEP-1998;	98US-0100849P.
PR	12-JUN-1998;	98US-0089105P.	PR	18-SEP-1998;	98US-0101014P.
PR	12-JUN-1998;	98US-0089112P.	PR	18-SEP-1998;	98US-0101068P.
PR	16-JUN-1998;	98US-0089514P.	PR	23-SEP-1998;	98US-0101471P.
PR	16-JUN-1998;	98US-0089538P.	PR	23-SEP-1998;	98US-0101472P.
PR	17-JUN-1998;	98US-0089653P.	PR	23-SEP-1998;	98US-0101475P.
PR	17-JUN-1998;	98US-0089653P.	PR	23-SEP-1998;	98US-0101477P.
PR	18-JUN-1998;	98US-0089952P.	PR	24-SEP-1998;	98US-0101738P.
PR	19-JUN-1998;	98US-0090246P.	PR	24-SEP-1998;	98US-0101739P.
PR	22-JUN-1998;	98US-0090252P.	PR	24-SEP-1998;	98US-0101743P.
PR	22-JUN-1998;	98US-0090254P.	PR	24-SEP-1998;	98US-0101922P.
PR	24-JUN-1998;	98US-0090429P.	PR	25-SEP-1998;	98US-0101786P.
PR	24-JUN-1998;	98US-0090435P.	PR	25-SEP-1998;	98US-0102207P.
PR	24-JUN-1998;	98US-0090444P.	PR	29-SEP-1998;	98US-0102240P.
PR	24-JUN-1998;	98US-0090461P.	PR	29-SEP-1998;	98US-0102330P.
PR	24-JUN-1998;	98US-0090535P.	PR	29-SEP-1998;	98US-0102331P.
PR	25-JUN-1998;	98US-0090540P.	PR	30-SEP-1998;	98US-0102487P.
PR	25-JUN-1998;	98US-0090676P.	PR	30-SEP-1998;	98US-0102570P.
PR	25-JUN-1998;	98US-0090678P.	PR	30-SEP-1998;	98US-0102571P.
PR	25-JUN-1998;	98US-0090688P.	PR	01-OCT-1998;	98US-0102684P.
PR	25-JUN-1998;	98US-0090690P.	PR	01-OCT-1998;	98US-0102687P.
PR	25-JUN-1998;	98US-0090694P.			
PR	25-JUN-1998;	98US-0090695P.			
PR	25-JUN-1998;	98US-0090696P.			
PR	26-JUN-1998;	98US-0090696P.			
PR	26-JUN-1998;	98US-0090863P.			
PR	26-JUN-1998;	98US-0090863P.			
PR	01-JUL-1998;	98US-0091010P.			
PR	01-JUL-1998;	98US-0091359P.			
PR	01-JUL-1998;	98US-0091544P.			
PR	02-JUL-1998;	98US-0091478P.			
PR	02-JUL-1998;	98US-0091486P.			
PR	02-JUL-1998;	98US-0091626P.			
PR	02-JUL-1998;	98US-0091628P.			
PR	02-JUL-1998;	98US-0091632P.			
PR	04-JUL-1998;	98US-0094006P.			
PR	04-AUG-1998;	98US-0095282P.			
PR	10-AUG-1998;	98US-0095998P.			
PR	10-AUG-1998;	98US-0096012P.			
PR	17-AUG-1998;	98US-0096757P.			
PR	17-AUG-1998;	98US-0096766P.			
PR	17-AUG-1998;	98US-0096867P.			
PR	17-AUG-1998;	98US-0096891P.			
PR	18-AUG-1998;	98US-0096949P.			
PR	18-AUG-1998;	98US-0096949P.			

Query Match 41.1%; Score 3101.5; DB 6; Length 1743;
Best Local Similarity 41.8%; Pred. No. 2,1e-130;
Matches 725; Conservative 0; Mismatches 421; Indels 587; Gaps 102;

QY	1	GATCTGGG-----GTGCTGCCAGGAAAAGCAAATCTTGGAGGTCGCGCTCT---CCCCGGTAT-CAG 102
Db	12	CGACAGGGCAGGAGAGGAACTTGCCTGCAAGA-CATATTTTGTCCAAAATGGCATC--- 67
QY	53	TGATTTTAA-----AATCTTGTCTGGGAGAGGCGCGCTCT---CCCCGGTAT-CAG 102
Db	68	TTACCTTTATGGAGTACTCTTTGCTGTTG-----GCCTCTGTGCTCCAATCTACTG 118
QY	103	CGCTTCTCTCATTCTTTGAATCC---CGGCTCCGGGTCTTGGGCTC--AGACGAGCGG 157
Db	119	TGTGTCCCGGCAATGCCCCAGTGCATACCCCGCTTCTCTCCACAAGAGACCCC- 177
QY	158	GAGGAGCC-----TGTTTCCAATTTAAGCGGCTGTG---AACGCCAGG--CGCGG 205
Db	178	----CTGCTCTCAGGTGATTCCTTCAACCGACTTTGCTTCCGCTATACCGCAGG 233
QY	206	CGGG----GGCGGGGCGGAGCGGCGCCATTT-----TGAATAAGAGCGGTGC 249
Db	234	CTGTTTGGAGACCCCGAGTCAAGACATCTTCTTCTCCCTGTGAGTGTCTCCACTTCC 293

QY 250 CT-TCCAGGC-----AGGCTCTAATAGTGACC-----GCC--- 278
Db 294 CTGGCCATGCTCTCCCTTGGGGCCCACTCAGTCACCAAGACCCAGATTCTCCAGGGCCNG 353
QY 279 -GGGGCAGCGGTGGCGC-----GT-TGCAGGTCACTGTAGCGGACTCTCTTTTGGTT 328
Db 354 GGGTTCAACTCACAACACACAGAGTCTGCCATCCACCAGGGCTTCCAGCACCTGGTT 413
QY 329 TTCTTTCTCTTTGG-----GGC-----ACCT-----CTGGAC-----TCACCTCCCGCAGC 367
Db 414 CACTCACTGACTGTTCACCAAGAACCTGACCTTGAAGATGGGAAGTGGCTCTTCGTC 473
QY 368 ATGAAGCGGTGAGCCGGTGC-GC-----GGC-TGCTACGAGGGGTGTGTCT 413
Db 474 AAGAAGGAGCTG---CAGCTGCAGGCAAAATTTCTTGGGCAATGTCAGAGGCTGTATGAA 530
QY 414 GC-----CTGT-----CGGA-----ACGAGCTGGCCATCGCCGGCGCGAGG---AAG 457
Db 531 GCAGAAGTCTTTCTCAGATTCTC-CAACCCCTCATTGCCCAGG--CGAGGATCAAC 587
QY 458 GGCCCGGCAGCTGAGGAGCGGCTGAGCTTGTCTGACGACATGAACCACTGC-----TACTC 513
Db 588 AGCCATGTGAAGAAAGAACCCAGGGAAGGTTGTAGACATAATCCAAGGCTTGACCTT 647
QY 514 CCGCTGCG--GGGAAGTGGT---ACCCGGAGTC-----CC-----GAGAGGCAC--T 553
Db 648 CTGACGCCATGTTCTGGTGAATCAATTTCTTTAAAGCCAAGTGGGAGAGGCCCTTT 707
QY 554 CAGCTT-----AG-----CCAGGTGG-----CAAGTGG-----AA 571
Db 708 CACCTTGAATATACAAGAAAGAACTTCCCATTTCTGGTGGCGGAGGAGTCACTGTGCAA 767
QY 572 ATCC-----TACAG--CGCGT-----CATCG-----ACTAC 595
Db 768 GTCCCATGATGACCAAGAAAGACAGTTCGCTTTTGGGTGGATACAGAGCTGAAGTGC 827
QY 596 ATTCTCAGCTGCAAGTAG--TCTGCGCGA-GCC--AGCC--CTGTGACC--CCCT--- 643
Db 828 TTT---GTGCTGCAAGATGATTACAGGAGATGCGGTGCGCTTCTTGTCTCCCTAGC 884
QY 644 -----GAT--GGCCCC-----CA--CTTCCCATCCAGACAGCGCA----- 675
Db 885 AAGGGCAAGATGAGGCAACTGGAACAGGCTTGTGACCCAGAACACTGATAAAGTGGAGC 944
QY 676 -GCTCGCTCGGAATCTG-----TCACTCCCAACGACAAAGAGACTTTTTCG 721
Db 945 CACTCACTCAGAAAAGGTGGATAGAGTGTTCATCCCC-----AGATTTC-C 991
QY 722 CA---CTGACTCGGCGGTGCTGACACCTCCAGAACGAGGTGCTGGCGCCGTTCTGC 778
Db 992 CATTTCTGCTCTCTAC-AACTGGAACCATCTCTCCGAGATG--GGCATCCAAATATGC 1048
QY 779 CTGGGAGC-----CCGG-----GAACCTCTCTGCGCG----- 806
Db 1049 CTTTGAACAAAATGCTGATTTTCTGGAATTGCAAGAGAGACTC-CCTGCAAGTTTCTA 1107
QY 807 ----AAGCGGACGGCAG-----GGATG-----GGCC----- 829
Db 1108 AAGCAACCCCAAGGCTGTGTGGATGTCACTGAAGAGGGCACTGAGGCCACAGCAGCTA 1167
QY 830 -----CGAATTC--GCCC-----TG--CCCACTTGAATTCACCAATCCCTT 868
Db 1168 CCACCAACCAAGTTATAGTCGATCGAAGGATGTCTCCCTCTT-ACCTCACTGTCTCTTC 1226
QY 869 CTGAGACTAAACCTGGTGT-----CA---GGAGCGAAGGACTGTGAAC----- 912
Db 1227 AATAGGACCT--TCTGATGATGATTACAAATAAGCCACAGAGGTATTCTCTTTCTAG 1284
QY 913 ----TGTAG-----CTGGAAGACCGAGAGTACTGTGCG-----ACGAGCTGGCGCAG 958
Db 1285 GGAAGTGGAAAATCCCACTAAATCTAGTGGGAAATGGCCTGTAACTGATGG----- 1339
QY 959 GTCAACCTGCT-----CCCAACCCCAACCC-----CAGATTCTTAAGG--TC 996

RESULT 13

ABU98260

ID ABU98260 standard; protein; 1743 AA.

XX AC ABU98260;

XX XX 30-JUL-2003 (first entry)

XX DE Novel human secreted and transmembrane protein #3.

XX KW Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy;

XX KW chondrocyte stimulator; tumour; adrenal tumour; lung tumour;

XX KW colon tumour; breast tumour; prostate tumour; rectal tumour;

XX KW cervical tumour; liver tumour; chromosome identification.

XX OS Homo sapiens.

XX XX US2003017544-A1.

XX XX 23-JAN-2003.

XX XX 21-JUN-2002; 2002US-00176915.

XX PR 18-SEP-1997; 97US-0059263P.

XX PR 18-SEP-1997; 97US-0059266P.

XX PR 17-OCT-1997; 97US-0062250P.

XX PR 21-OCT-1997; 97US-0063486P.

XX PR 24-OCT-1997; 97US-0063120P.

XX PR 24-OCT-1997; 97US-0063121P.

XX PR 28-OCT-1997; 97US-0063540P.

XX PR 28-OCT-1997; 97US-0063541P.

XX PR 28-OCT-1997; 97US-0063564P.

XX PR 29-OCT-1997; 97US-0063734P.

XX PR 31-OCT-1997; 97US-0063870P.

XX PR 31-OCT-1997; 97US-0064103P.

XX PR 13-NOV-1997; 97US-0065311P.

XX PR 21-NOV-1997; 97US-0066120P.

XX PR 24-NOV-1997; 97US-0066772P.

XX PR 11-DEC-1997; 97US-0069335P.

XX PR 12-DEC-1997; 97US-0069425P.

XX PR 17-DEC-1997; 97US-0069870P.

XX PR 18-DEC-1997; 97US-0068017P.

XX PR 10-MAR-1998; 98US-0077450P.

XX PR 11-MAR-1998; 98US-0077632P.

XX PR 11-MAR-1998; 98US-0077649P.

XX PR 20-MAR-1998; 98US-0078866P.

XX PR 20-MAR-1998; 98US-0078939P.

PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090245P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090576P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-0090699P.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 10-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096157P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-0101751P.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.

PR	30-SEP-1998;	98US-0102570P.	
PR	30-SEP-1998;	98US-0102571P.	
PR	01-OCT-1998;	98US-0102684P.	
PR	01-OCT-1998;	98US-0102687P.	
PR	02-OCT-1998;	98US-0102965P.	
PR	06-OCT-1998;	98US-0103258P.	
PR	06-OCT-1998;	98US-0103449P.	
Query Match 41.1%; Score 3101.5; DB 6; Length 1743; Best Local Similarity 41.8%; Pred. No. 2.1e-130; Matches 725; Conservative 0; Mismatches 421; Indels 587; Gaps 102;			
QY	1	GATCTGGG-----GTGCTGCCAGGAAAGCAAAATCTCGAAGTTAATGTTTGGAG 52	
DB	12	GGACAGGGCAGGAGAGAGAAACCTGCNAAGA-CATAATTTTGTTCNAATGGCATC--- 67	
QY	53	TGATTTTAA-----AATCTTGCTGGCGAGAGGCCGCTCT---CCCCGGTAT-CAG 102	
DB	68	TTACCTTTATGGAGTACTCTTTGCTGTTG-----GCTCTGTGCTCCATCTACTG 118	
QY	103	CGCTTCCTCAATCTTTGAATCC---CGGGTCGCGGTCTTCGGGTC---AGACAGCGG 157	
DB	119	TGTGTCGGCGCAATGCCCCAGTCATACCCCGCCCTTCTCCACAAAGAGCACCC- 177	
QY	158	GAGGAAGCC-----TGTTTCAATTTAAGGGGCTGTG---AAGCCCAAG---GCCGG 205	
DB	178	-----CTGCTTCACAGGTGTATCCCTCAACCGACTTTGCTTCGGCTATACCGCAGG 233	
QY	206	CGGG-----GGCGGGCGGAGCGGGCCATTT-----TGAATAAAGAGCGGTGC 249	
DB	234	CTGGTTTGGAGAGCCCGAGTCAGAAATCTTCTTCTCCCTGTGTAGTCTCTCCACTTCC 293	
QY	250	CT-TCCAGGC-----AGGCTTATAAGTGACC-----GGC--- 278	
DB	294	CTGGCCATGTCTCCCTTGGGCGCCACTCAGTCACCAAGACCCAGANTTCTCCAGGCGCTG 353	
QY	279	-GGCGGAGCGTGGCGG-----GT-TGCAGGTCACTGTAGCGGACTCTTTTGGTT 328	
DB	354	GGCTTCAACCTCACACACACAGAGTCTGCCATCCACAGGGCTTCCAGCACCTGGTT 413	
QY	329	TTCTTTCTTTGG-----GGC---ACCT---CTGGAC-----TCATCTCCCGACG 367	
DB	414	CACCTCACTGACTGTTCACGAAAGACCTGACCTTGAAGATGGGAAGTCCCTCTTTGCTC 473	
QY	368	ATGAAGCGCTGAGCGCGTGC---GC-----GGC---TGCTACGAGGGGCTGTGCT 413	
DB	474	AAGAGAGAGCTG---CAGCTGCAGGCAATTTCTTGGCAATGTCAAGAGGCTGTATGAA 530	
QY	414	GC-----CTGT-----CGGA-----ACGAGCTGGCCATGCCCCGGGCGGAGG--AAG 457	
DB	531	GCAGAGCTCTTTCTACAGATTTCTC-CACCCCTCCATTTGCCAGG--CGAGGATCAAC 587	
QY	458	GGCCCGCAGCTGAGAGCGCTGAGCTTCTGCGACGACATGAACCACTGC---TACTC 513	
DB	588	AGCAATGTAAAGAGAGACCCAGGGAAGTTGTAGACATATCCAGGCTTGACCTT 647	
QY	514	CCGCGCTGC--GGGAAGTGGT---ACCGGAGTC-----CC-----GAGAGGCAAC--T 553	
DB	648	CTGACGCCCATGTTCTGTGTAATCAATTTCTTTAAAGCCAAAGTGGGAGAGGCCCTTT 707	
QY	554	CAGCTT-----AG-----CCAGTGG-----AA 571	
DB	708	CACCTTGATATACAGAAAGAACTTCCCATTCCTGTGGCGAGCAGGTCACTGTGCA 767	
QY	572	ATCC-----TACAG--CGCGT-----CATCG-----ACTAC 595	
DB	768	GTCCCCATGATGCACCAAGAGACAGTTCGCTTTTGGGTGGATACAGAGCTGAAGTGC 827	
QY	596	ATTCTGACCTGAGGTAG---TCCTGGCGGA-GCC--AGCC--CCTGGACC-CCCT--- 643	
DB	828	TTT---GTGCTGCAGATGGATTACAGGGAGATGCCGTGGCCCTTCTTTGTCTCCCTAGC 884	
QY	644	-----GAT--GGCCCC-----CA--CCTTCCCATCCAGACAGCCGA----- 675	

DB	885	AAGGCAAGATGAGGCAACTGGAAACAGGCCCTTCTCAGCCAGAAACACTGATAAAGTGAGC 944	
QY	676	-GCTCGCTCCGGAACCTTG-----TCATCTCCAAACGACAAAAGAGGCTTTTGC 721	
DB	945	CATCACTCCAGAAAGGTGGAAGAGTGTTCATCCCC-----AGATTTT-C 991	
QY	722	CA---CTGACTCGGCGCTGCTGACACCTCCAGAGCGAGGTGCTGGCGGCCCTTCTGC 778	
DB	992	CATTTCTGCTCTCTAC-AAATCTGGAACCACTCTCCGGAAGATG--GSCATCCAAAATGC 1048	
QY	779	CTGGGACC-----CCGG-----GAACCTCTCTCTCCGCG-- 806	
DB	1049	CTTTGACAAAATGCTGATTTTCTGGAATTCAGAGAGAGATC-CTGAGGTTCTTA 1107	
QY	807	----AAGCGGAGCGGAG-----GGATG-----GGCC----- 829	
DB	1108	AAGCAACCCACAGGCTGTGTGATGTCAAGAGAGGCACTGAGGCGCACAGCAGCTA 1167	
QY	830	-----CCAACTTC--GCCC-----TG--CCCACTTGACTTCCCAATCCCTT 868	
DB	1168	CCACCACTGATTCATAGTCGATCGAAGAGATGTCCTCTT-ACITCACTGTCTCTTC 1226	
QY	869	CCTGGAGACTAAACCTGGTGT-----CA---GAGCGAAGGACTGTGAAT----- 912	
DB	1227	AATAGGACCT--TCCTGATGATTAATAAAGCCACAGACGGTATTTCTCTTTCTAG 1284	
QY	913	----TGTAG-----CCTGAGAGCCAGAGTAGTCTGGCC-----ACCAGCTGGGCGAC 958	
DB	1285	GGAAGTGGAAATCCCACTAAATCTTAGTGGGAAATGGCCTGTTAACTGATGG----- 1339	
QY	959	GTCAACCTGTCT-----CCCAACCCACCCG-----CAAGTCTTAAG--TC 996	
DB	1340	--CACATGCTAATGCAAGAAATACAAACCATCTCTTCTTCTGTCTGAGGCTG 1397	
QY	997	TTTTTCAGAGC-GTGGAGGTG-----TGAAGAGGTGGCT-----GC-----TCTCC 1036	
DB	1398	ATTTGACCCAGTGGAGTGGATTCGCTGGCAGGATGCCACTTCCAGGCTCAATCACC 1457	
QY	1037	AAACTAT-----GCCAAGGC-----GGC-----GGCAGAG--CTG 1064	
DB	1458	AAACATCAACAGGAGCCCGAGTCACAGCCACACACCCATTAACCCAGTCAGTGCCTT 1517	
QY	1065	GTCTTCTGG-TCTCTTGG-AGAAAGTTC-----TGTTCG-----CCTGATTT--A 1107	
DB	1518	TTCCACAAATTTCTCCAGGTAACCTAGCTTCATGGGATGTGCTGGGTATCCCATATTTCCA 1577	
QY	1108	T-----GAACTCTATATAGAG-----TATATAGTTTGTACCTTTTTCACAGG 1152	
DB	1578	TTCTTGGGGCTCCAGGATGGAAATACGCCAACCCAGGTTAGGCACCTCTATTGAGA 1637	
QY	1153	----AAGGTGACTTTCTGTAACAATCGATGTATATTAACCTTTTATAAAA 1200	
DB	1638	ATTACATTAACACATTCATTAATAAACTAAATATGAATTCAAAAAATAAAAAA 1690	

RESULT 14
 ABU91966
 ID ABU91966 standard; protein; 1743 AA.
 XX
 AC ABU91966;
 XX
 DT 11-AUG-2003 (first entry)
 XX
 DE Novel human secreted and transmembrane PRO protein #3.
 XX
 KW Human; gene therapy; chromosome identification; tissue typing.
 XX
 OS Homo sapiens.
 XX
 PN US2003027277-A1.
 XX
 PD 06-FEB-2003.

Db 1458 AAACATCAACAGGGACCCAGTCACACCAACCAACCCATTAAACCCAGTCAGTCCCTT 1517
QY 1065 GTCTTCTGG-TCTCCTTGG-AGAAAGGTTT-----TGTTGC-----CCTGATTT--A 1107
Db 1518 TTCCACAATTTCCACAGGTAACCTTCATGGGATGTTGCTGGTTACCATATTCCCA 1577
QY 1108 T-----GAATCTATATATAGAG-----TATATAGTTTGTACTTTTACAGG 1152
Db 1578 TTCTTTGGGCTCCAGGAATGAATAGCCACCCAGTTAGGACCTCTATTGCAGA 1637
QY 1153 -----AAGCTGACTTCTGTAAACAATGCGATGTATATTAACCTTTTATAAAA 1200
Db 1638 ATTACATAACACATTCATATAAATGAATATGAATTCAAAAAATAAAAAA 1690

RESULT 15
ABU85270
ID ABU85270 standard; protein; 1743 AA.
XX AC ABU85270;
XX DT 30-JUN-2003 (first entry)
XX DE Novel human secreted and transmembrane protein #3.
XX KW Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy;
XX KW chondrocyte stimulator; chromosome mapping; gene mapping;
XX KW transgenic animal; knock-out animal; tumour.
XX OS Homo sapiens.
XX PN US2003032114-A1.
XX PD 13-FEB-2003.
XX PF 20-JUN-2002; 2002US-00176919.
XX 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081045P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.

PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089539P.
PR 17-JUN-1998; 98US-0089633P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090689P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.

PR	26-JUN-1998;	98US-0090863P.	Best Local Similarity	41.8%;	Pred. No. 2.le-130;	
PR	28-JUN-1998;	98US-0091010P.	Matches	725;	Conservative	0;
PR	01-JUL-1998;	98US-0091359P.			Mismatches	421;
PR	01-JUL-1998;	98US-0091544P.			Indels	587;
PR	02-JUL-1998;	98US-0091478P.			Gaps	102;
PR	02-JUL-1998;	98US-0091486P.	QY	1	GATCTGGG-----GTGCTGCCAGGAAAGCAAAATCTTGGAAAGTAAATGGTTTGGAG	52
PR	02-JUL-1998;	98US-0091626P.	Db	12	GCACAGGCGCAGGAGAGAGAACTGCAAGA-CATATTTTGTTCCTCCAAAATGGCATC---	67
PR	02-JUL-1998;	98US-0091628P.	QY	53	TGATTTTAA-----AATCCTTGTGGCGAGAGGCGCGCTCT---CCCCGGTAT-CAG	102
PR	02-JUL-1998;	98US-0091632P.	Db	68	TTACCTTTATGGAGTACTCTTTGTGTG-----GCCTCTGTGCTCAATCTACTG	118
PR	04-AUG-1998;	98US-0095282P.	QY	103	CGCTTCCTCATTTTGAATCC---GGGGTCCGCGGTCTTTCGGCGTC---AGACAGCGG	157
PR	10-AUG-1998;	98US-0096012P.	Db	119	TGTGTCCCGCGCAATGCCCGAGTGCATACCCCGCGCTTCTCTCCAAAGAGACACC-	177
PR	17-AUG-1998;	98US-0096757P.	QY	158	GAGGAAGCC-----TGTTTTCATTTAAGCGGGCTGTG---AACGCCAGG--GCCGG	205
PR	17-AUG-1998;	98US-0096867P.	Db	178	-----CTGCTTCACAGGTGTATTCCTCAACAGCTTTGCCCTTCGCGCTATACCGAGG	233
PR	17-AUG-1998;	98US-0096897P.	QY	206	CGGG-----GGCGGGCGAGGCGGGCCATTT-----TGAATAAAGAGCGGTGC	249
PR	18-AUG-1998;	98US-0096949P.	Db	234	CTGGTTTTTGGAGACCCCGAGTCAGAACATCTTCTTCCCGCTGTGAGTGTCTCCACTCC	293
PR	26-AUG-1998;	98US-0097954P.	QY	250	CT-TCCAGGC-----AGGCTCTATAAGTGACC-----GCC--	278
PR	26-AUG-1998;	98US-0097971P.	Db	294	CTGGCCATGCTCTCCCTTGGGGGCCCACTCAGTCACCAAGACCCAGATTTCCAGGGCGTG	353
PR	01-SEP-1998;	98US-0098716P.	QY	279	-GGCGGAGCGTGCGCGC-----GT-TGCAGGTCACTGTAGCGGACTTCTTTTGGTT	328
PR	01-SEP-1998;	98US-0098723P.	Db	354	GGCTTCAACCTCACACACACAGAGTCTGCCATCCACCGAGGCTTCCAGCACTGGTT	413
PR	02-SEP-1998;	98US-0098803P.	QY	329	TTCTTCTCTTTGG-----GGC-----ACCT---CTGGAC-----TCACTCCACAGC	367
PR	02-SEP-1998;	98US-0098821P.	Db	414	CACTCACTGACTGTTCCCGCAAGACCTGACCTGAAGATGGAAAGTGCCCTCTTGGTC	473
PR	02-SEP-1998;	98US-0098843P.	QY	368	ATGAAGCGCTGAGCCCGGTGC--GC-----GGC--TGCTACGAGCGGTGTGCT	413
PR	09-SEP-1998;	98US-0099602P.	Db	474	AAGAGGAGCTG---CAGCTGAGGCAAAATTTCTTGGCAATGTCAAGAGCTGTATGAA	530
PR	10-SEP-1998;	98US-0099741P.	QY	414	GC-----CTGT-----CGGA---ACGCAGTCTGGCCATGCGCCGGGCGCGAGG--AAG	457
PR	10-SEP-1998;	98US-0099754P.	Db	531	GCAGAAAGTCTTTCTACAGATTTCTC-CACCCCTCCATTTGCCAGG--CGAGGATCAAC	587
PR	10-SEP-1998;	98US-0099763P.	QY	458	GGCCCGGCGAGCTGAGGAGCGCTGAGCTTGTGCGACACATGAACCACTGC---TACTC	513
PR	15-SEP-1998;	98US-0100388P.	Db	588	AGCCATGTGAAAAGAGAGACCCAGGGAAGTTGTAGACATATCCAAAGGCTTGACCTT	647
PR	16-SEP-1998;	98US-0100662P.	QY	514	CCGCTGC--GGGAACTGGT---ACCCGGAGTC-----CC-----GAGAGGAC--T	553
PR	16-SEP-1998;	98US-0100664P.	Db	648	CTGACGGCCATGGTTCTGTGTAATCACAATTTCTTTAAAGCCAAAGTGGGAGAGCCCTTT	707
PR	16-SEP-1998;	98US-0101751P.	QY	554	CAGCTT-----AG-----CCAGGTGG-----AA	571
PR	16-SEP-1998;	98US-0101751P.	Db	708	CACCTTGAATATACAAGAAAGAACTTCCATTTCTGTGGCGAGAGGTCACGTGTCAA	767
PR	16-SEP-1998;	98US-0101933P.	QY	572	ATCC-----TACAG--CGCGT-----CATCG-----ACTAC	595
PR	17-SEP-1998;	98US-0101684P.	Db	768	GTCCCATGATGCACCAAGAAAGACAGTTCCTTTTGGGTGGATACAGAGCTGAAGTGC	827
PR	17-SEP-1998;	98US-0101684P.	QY	596	ATTCTGACCTGCAGGTAG---TCTGCGCA-GCC--AGCC--CCTGGACC-CCCT---	643
PR	23-SEP-1998;	98US-0101471P.	Db	828	TTT---GTGCTGCAGATGGATTACAAGGGAGATCCGCTGGCTTCTTTGTCTCTCCCTAGC	884
PR	23-SEP-1998;	98US-0101475P.	QY	644	-----GAT--GGCCCC-----CA--CCTTCCCATCCAGACAGCCGA-----	675
PR	23-SEP-1998;	98US-0101477P.	Db	885	AAGGCAAGATGAGGCAACTGGAACAGGCTTGTGACCGAGAACACTGATAAAGTGAGC	944
PR	24-SEP-1998;	98US-0102571P.	QY	676	-GCTGCTCCGGAAGTCTG-----TATCTCCAAACGACAAAGAGGCTTTTGC	721
PR	24-SEP-1998;	98US-0102684P.	Db	945	CACCTCACTCCAGAAAAGGTGGATAGAGGTGTTTCATCCCC-----AGATTTT-C	991
PR	24-SEP-1998;	98US-0102687P.	QY	722	CA---CTGACTCGCGGTGTCTTACACACCTCCAGAACGCGAGGTGCTGCGGCCCTTCTGC	778

Query Match 41.1%; Score 3101.5; DB 6; Length 1743;

Db 992 CATTCTGCTCTCTAC-AATCTGGAAACCATCTCTCCGAGATG--GGATCCAAATGC 1048
Qy 779 CTGGGACC-----CCGG-----GAACCTCTCTGCGG----- 806
Db 1049 CTTTGACAAATGCTGATTTTCTGGAATTGCAAGAGAGACTC-CCTGCAGTTTCTA 1107
Qy 807 ----AAGCCGGACGGCAG----GGATG-----GGCC----- 829
Db 1108 AAGCAACCCACAAGGCTGTGCTGATGTAGTGAAGGGCACTGAGGCCACAGAGTA 1167
Qy 830 -----CCAACTTC---GCC-----TG--CCCACTTGACTTACCAAATCCCTT 868
Db 1168 CCACCACCAAGTTTCATAGTCGATCGAAGATGGTCCCTCTT-ACCTTCACTGTCTCCTC 1226
Qy 869 CTTGGAGACTAAACCTGTGCT-----CA---GGAGGAGGACTGTGAAC----- 912
Db 1227 AATAGGACCT--TCCTGATGATTAACAAATAAGCCACAGACGTATTCTCTTTCTAG 1284
Qy 913 ----TGTAG-----CCTGAAGGCCAGAGCTAGCTCTGSCC-----ACCAGCTGGGCGAC 958
Db 1285 GGAAGTGGAAATCCCACTAAATCCTAGTGGGAANTGGCCGTGTTAACTGATG----- 1339
Qy 959 GTCACCCTGCT-----CCACCCCAACCCG-----CAAGTTCTAAGS--TC 996
Db 1340 --CACATTGCTAATGCACAAGAAATAACAAACACATCCCTCTTCTGTTCTGAGGGTGC 1397
Qy 997 TTTTCAGAGC-GTGGAGTG-----TGAAGGAGTGGCT-----GC-----TCTCC 1036
Db 1398 ATTTGACCCAGTGGAGCTGGATTGCTGACAGGGATGCCACTTCCAAGGCTCAATCACC 1457
Qy 1037 AAACATAT-----GCCAAGGC-----GGC-----GGCAGAG--CTG 1064
Db 1458 AAACCATCAACAGGGACCCAGTCACAAGGCCAACACCCATTAAACCCAGTCAGTGCCTT 1517
Qy 1065 GTCTTCTGG-TCTCCTTGG-AGAAAGTTC-----TGTTC-----CCTGATTT--A 1107
Db 1518 TTCCACAAATTCTCCCAAGGTAACTAGCTTCATGGGATGTTGCTGGGTACCATATTTCCA 1577
Qy 1108 T-----GAACTCTATAATAGAG-----TATATAGGTTTGTACCTTTTITACAGG 1152
Db 1578 TTCCTTGGGGTCCCAAGGAATGGAATACGCCAACCCAGGTAGGCACCTCTATTTCGAGA 1637
Qy 1153 -----AAGGTGACTTTCTGTAAACATGCGATGTATATTAACCTTTTATAAAA 1200
Db 1638 ATTACAATAACACATTCAATAAAACTAAATATGAATTCAAAAAAATAAAAAA 1690

Search completed: May 7, 2004, 15:08:41
Job time : 102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:06:59 ; Search time 25 Seconds
(without alignments)
2484.242 Million cell updates/sec

Title: X59111
Perfect score: 7544
Sequence: 1 GATCTGGGGTGGCCAGCA.....TTAAACTTTTATAAAGTT 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000.

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

ALIGNMENTS

RESULT 1
US-09-738-884-1

; Sequence 1, Application US/09738884
; Patent No. 6391606
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00849
; CURRENT APPLICATION NUMBER: US/09/738.884
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2211
; TYPE: PRT
; ORGANISM: Human
US-09-738-884-1

Query Match 45.7%; Score 3447.5; DB 4; Length 2211;
Best Local Similarity 36.9%; Pred. No. 9e-153;
Matches 809; Conservative 0; Mismatches 355; Indels 1027; Gaps 124;

QY	4	CTGGGGTGTGCTCCAGGAAAAAGCAAAATCTGGAAGTTAA--TGGTTTT--GAGTCAATTT	59
Db	37	CTGGGGGGTCC--GGCTCCCAAGATCCGCTCGCACGCTGGCACAAGGAGCGGCTGT	94
QY	60	TAAATCCTTGTCTGGCGGAGGCGC-----CGCCTCTCCCG-----	95
Db	95	ACCGCG--TGCAGGAGGA-CGGCCTGAGGCTGTGGTTCAGCGCGGCATCCGCGTGGC	151
QY	96	-GTATCAGCGCTTCTCAT-----TCTTTGA-----ATCCGC---GGC-----TCGG	133
Db	152	CATCGCAGCACATCTTCTTCGTGACACATCGAGCGGTCCGCGAGGCGCACCACTCG	211
QY	134	C-GGTCTTGGGG--TCAG---ACC---ACCGGAG-GRAGCCT---GTTTGAATTT	178
Db	212	AGGGCGTCTGGCGGCTTCGGGGGTCCTTCGCGCAGCGCTGCTCTCACCATCGCTTCA	271
QY	179	AAG-----CGGG--CTGTGAACGCCCCAGGCG-----CGGC---GGG	209
Db	272	AGGGCGCGCGCAAGAACCTGGACCTTGGCGGCGCCCAAGGAGCGGAGCGGTGGG	331
QY	210	GGCGGGG-----CCGAG-GCG---GGCCATTTTGAATAA---AGAGCGGTG	248
Db	332	TGCGCGTCTGACCAAGCTCCGCGCGCTCGACGCCATGACCGCGCGAGCGCGCTAG	391
QY	249	-----CCT-----TCAGGCGAGGCTCTATAAGTGACCGC-----	277

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3447.5	45.7	2211	4	US-09-738-884-1
2	3295.5	43.7	1417	4	US-08-900-230-3
3	3123.5	41.4	2088	4	US-09-548-372D-13
4	3123.5	41.4	2088	4	US-09-548-367D-13
5	3123.5	41.4	2088	4	US-09-551-853D-13
6	3105	41.2	2601	4	US-09-627-650B-9
7	3105	41.2	2601	4	US-09-436-063C-9
8	3097.5	41.1	2508	4	US-09-627-650B-7
9	3097.5	41.1	2508	4	US-09-436-063C-7
10	3097.5	41.1	2544	4	US-09-627-650B-3
11	3097.5	41.1	2544	4	US-09-436-063C-3
12	3073.5	40.7	1345	2	US-08-977-767-3
13	2971.5	39.4	1917	4	US-09-627-650B-5
14	2971.5	39.4	1917	4	US-09-436-063C-5
15	2954.5	39.2	1652	4	US-09-627-650B-1
16	2954.5	39.2	1652	4	US-09-436-063C-1
17	2913	38.6	1400	4	US-08-630-915A-37
18	2874	38.1	1497	4	US-09-060-854B-2
19	2552	33.8	1128	4	US-09-627-650B-11
20	2552	33.8	1128	4	US-09-436-063C-11
21	2331	30.9	1388	4	US-09-463-048A-6
22	2050	27.2	908	4	US-08-714-741-44
23	1812	21.4	717	4	US-07-644-460-37
24	1391.5	18.4	801	1	US-07-906-349A-6
25	1378.5	18.3	605	4	US-09-976-594-616
26	1269	16.8	1184	4	US-10-153-064-89
27	1267	16.8	1332	1	US-07-609-716-41


```

; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52241-C/JPW/ADM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-900-230-3

Query Match      43.7%; Score 3295.5; DB 4; Length 1417;
Best Local Similarity 48.6%; Pred. No. 7e-146f;
Matches 690; Conservative 0; Mismatches 412; Indels 317; Gaps 101

QY    1   GATCTGGGGTGTGCTG-----CCAGGAAGAAGCA---AATTCTGGAAGTAATGTTTGAGT 53
DB    81   GGTGCGGGAGTTAGATCCGGGGTCACAGCAACAGAACTGGGGGCTTGT--CCTGAG- 137
QY    54   GATT-----TTTAATTCCTTG-----CTGGCGGAGAG-----GCCC---GCCT 88
DB   138   GATTCACAGTCTCTCTTCCCAGGTGCCGTCTGATGGGGAGATGGCTGATGCCACAGAACAT 197
QY    89   CTCGCCGGTATCAGCGCTTCCTCATCTCT-TTGAATCCGGCGCT----CGCGGTCCTTCGG 143
DB   198   TTCACTGG--ACAGC---CCAGGGAGTGTGGGGCCGTGCGAGTGCCTGTGGTCTTTGC 251
QY   144   CQTCAACAAGCCGGAGAAAGCTGTTTGCMAAT-----TTAAGCGGGCTGTGAAC----- 193
DB   252   C--CTAATCTTCTCTGGGCACAGTAGTGGGCAATGGGCTGGTGTGGCAGTGCTCTCTGCGAG 309
QY   194   ---GCCC-AGGGCC-GCGGGGGGGGGGC--CGAGCGGGGCCATTTTGAATAAAGAGGC 245
DB   310   CTGTGCCCGAGTGCCTGTGCAGAGAGCTGGCAGACCACAGCACTGTT-----C 357
QY   246   GTGC-CTTCCAAGC--AGGCTCTTAAGTGTACGCCCGCGCGAGC-GTGC-GCG---CGT 297
DB   358   ATCCTCAACCTGGCGTGGC-----TGACCTCTGCTTCACTCTGTGCTGTGCCCT 409
QY   298   TGCAGTGTCACTAGCGGACATCTTT-TTGGTTTTCTTTCTTTGGGGCAC-C-TCGTGA 354
DB   410   TCAGGCCCACTAT-----CTACAGCTGSGATGCTGGCTCTTTTGGGGCCCTCGTCTGCA 463
QY   355   -----CTCA-CTCCC-CAGCATGAAGGGCGCTG-AGC---CGGCTGCGCGGC 394
DB   464   AGSCCGTGCACCTGCTCATCTACCTACCATGATAGC-CAGACGTTTACGCTG-GCTCG 521
QY   395   TG-CTACAGAGG-CGGTGTGCTGCTGT-CGGAA--CGAGT-CTGGCCATCGCCCGGGGC 448
DB   522   TGTCTCCGTGACAGTACTCTGGCCGTGCGCACCCGCTGGGCTGCG--CGCGCCCTGCG 579
QY   449   -CG-AGGGAAGGGCCCG-GCAGCTCAGAGAGCGGTGTGCTG-----G 491
DB   580   ACCCGCGTTAACGCCCGCGCCGCACTGGGGCTGTGTGGCTGTCTGGCGGCGCTCTTCTG 639
QY   492   ACGACATGAACCACTGCTACTCCCC-----CTGCGG-GAACTGGTACC CGGAGT-C 541
DB   640   GCGCCCT--ACCTCAGCTACTACGSCACCGTGGCTACCGCGCGCTGGAGCTCTGGCGTG 697
QY   542   CCG-----AGAGCACTCAGCTTACCGAGGTGGAAT--CCTACAGCGCGTCACTGCAG 592
DB   698   CCGCCCTGGAGGACGCGCGCGCCCGGCC---TGACGTGGCCACCTTTCGCTGC--CGGC 753
QY   593   TACAATCTCCACC--TGC--AGGTAGTCTCTGSC---CGAGCCAGGCC-CTGAGCCCTCTG 644

```

US-09-548-372D-13

Query Match 41.4%; Score 3123.5; DB 4; Length 2088;
Best Local Similarity 37.5%; Pred. No. 8.5e-138;
Matches 780; Conservative 0; Mismatches 403; Indels 895; Gaps 126;

QY 1 GATCTGGGTGCTGCAGAGAAAGAAATTCGAGT-----TAATGTTTTGAG-- 52
DB 21 GCTCTGCTGGCGCTGGACCGCTCGCGCTGGAGGTACCCACTGATGTTAAAGTGG 80
QY 53 --TGATTTTAAATCC-----TTGC-----TGCGAGAG-----GCC--CCCTCT 90
DB 81 CCTGCTGGTGAACCCAGATTTGCCATGTTCTGTGCGAGCTGAACATGACATGAATGT 140
QY 91 CC-----CCGGT--ATCAGG--CTTCTCATTTCTTTGAATCGCGG-- 128
DB 141 CCAGAAATGGAGTGGGATTCAGATCCATCAGGGACCAAACTGCAATGATACCAAGGA 200
QY 129 ----CTCC--GCGTCTT-----CGGCTCAGACAG--CCGAGGAAGCTGTTT--CA 174
DB 201 AGGCATCTCGAGTATTCGAGAGTCTACCTGAATGCGAGATCAACCAATGTGTAGA 260
QY 175 ATTAAAGCGGTGTGAAGCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 226
DB 261 AGCCAAACCAACAGTACCATCCAGAACTGCTGCAAGCGCGCGCGCGCGCGCGCGCGCG 320
QY 227 ----CCATTTG-----AATAAAGAGC-----GTGCC-- 250
DB 321 CCATCCCATCTTGTGATCCCTACCGCTGCTTAGTTGTTGAGTGTGATGCTCT 380
QY 251 ----TTCCAGGAGG-----CTCTATAAGTG----- 272
DB 381 TCTCGTTCTTCAAGTGCATAATTTCTTACACAGGAGAGTGTGATTTTGGAAACTCA 440
QY 273 ----ACCGCGC-----GCGAGC--GTGCG-----CGCGTTGAGGT 304
DB 441 TCTTCACTGCGCACACCGTGCAGAGAGAGATGCTGAGTGAAGAGTACCACTTGCATGA 500
QY 305 C-AC---TGAGCGAC-----TTC-----TTTGTGTTT- 330
DB 501 CTACGGCATGTTGCTGCTGCTGCGAAATGCAAGTTCCGAGGGGTAGAGTTTGTGTGTTG 560
QY 331 ----CTTTCT-----CTTTGGGCACT-CTG-----GATC-- 357
DB 561 CCCACTGGCTGAAGAAAGTGAATGTGGAATCTGCTGATGCGGAGAGAGTACTCGGA 620
QY 358 ----ACTC---CCAGCA-----TGAAGGC----- 375
DB 621 TGTCTGTGGGCGGAGCAGACAGACTATGAGATGGAGTGAAGACAAAGTAGTAGA 680
QY 376 ----GCTGAG-----CCCGGTG-----CCCGGC 394
DB 681 AGTAGCAGAGAGGAAGTGTGAGGTGGAAGAGAAAGAGAGAGAGATGATGACGAGGA 740
QY 395 TGCT-----ACGAGCGGTG-----TGCTG-----CTGTGCG----- 421
DB 741 CGATCAGATGTTGATGAGGTAGAGGAGAGGTGAGGACCTTACGAAGAGCCACAGA 800
QY 422 --GAAC--GCAGTGTGCGCATGCGC-----CGGGCCGAGGAGAGG----- 459
DB 801 GAGAACCAACAGCATTTGCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 960
QY 460 ----CC--CGGCGCTG--AGGAGCGGTGA--GC--TTG-----CTGACGAC 496
DB 861 TCGAGTCTCTACACAGAGCGAGTACCTGATGCGGTGTGACAAAGTACTCGAGACAC 920
QY 497 ----ATGAACCACTGTACTCCCGCTGC-----GGGAACCTGTGTA 532
DB 921 TGGGGATGAGAAATGAACATGCCATTTCCAGAAAGCCAAAGAGAGAGGCTTGAGCCCAAGCA 980
QY 533 CCCGGA-----GTCCC-----GAGAGGCA----- 551
DB 981 CCGAGAGAGAAATGTTCCAGGTCTGAGAGAAATGGGAAGAGGCGAGAACGTCACCAAGAA 1040

QY 552 -----CT-----CAGCTTAGCCAGGTG-----GAAATCCT- 576
DB 1041 CTTGCTTAAAGCTGATTAAGAGCAGTATCCAGCATTTTCCAGGAGAAAGTGAATCTTT 1100
QY 577 -----ACAGC-----GGTTCATCGACT-----ACATTTCTCGACCTGC-AGGT 612
DB 1101 GBAACAGGAGACGCAACGAGAGACAGAGCTGTGGAGACACACATGGCCAGAGTGA 1160
QY 613 AGTCCTG-----GCGG--AGCAGAGCCCTGGA-----CCCC--CTGATGGC 649
DB 1161 AGCCATGCTCAATGAGCGCGCGCTGCGCTGCGAGAACTACATCACGCTCTGAGGC 1220
QY 650 --CCCCACCTTCCCATCCAGACAGCCGAGCTC-----GC-----TCCG-----G 686
DB 1221 TGTTCTCTCTCGGCTC--GTCA--CGTGTCAATATGCTAAAGAGATGTCGCGCAG 1276
QY 687 AACT-----TGTATCT--CCACGACAAAGAGCTTTTTC----- 721
DB 1277 AACAGAGGACAGACAGACACACCTTAAGCAITTCGAGCATGTGCGCATGTGCGATCCCA 1336
QY 722 -----CACTGACTCGG--CCGTGCTCTG--ACACCTCC-----AGAACGCA 758
DB 1337 AGAAGCCGTCTAGATCCGTCGAGGTATGACACACCTCCGTGTGATTTATGAGGCA 1396
QY 759 -GGTGTGGCGCGCTTCTGC-CTGGGACCCCGGAACTT-CTCTGCGG-----GAAGCC 811
DB 1397 TGAATCAGTCTCTCTCCTCTCTCT--ACAACGTG--CCTGCACTGCGCGAGGAGATTCA 1451
QY 812 GGACGGCAGGATGGG-----CCCAACT-----TCGCCCT-GGCCA 847
DB 1452 GGATGAAGTTGATGAGCTGCTTCAGAAAGAGCAAACTATTCAATGAGCTCTTGSCCA 1511
QY 848 CTTGACTT---CACCAA--ATCCCTTCTCGAGACTAAAC-CTGGTGC-----TCAG 893
DB 1512 CATGATTAGTGAACCAAGGATCAGTTAC-GGAAACGATGCTCTCATGCCATCTTTGACCG 1570
QY 894 GAGCGAAGG--ACTGTGAA-----CTT-----GT--AGCTGAA-----G 924
DB 1571 AAACGAAACACCGTGGAGCTCTTCCCGTGAATGAGAGTTTCACTGAGCTGACATCTCC 1630
QY 925 AGCC-----AGAGCTAGCTCTG-GCCA--CCAGCTGGG-----CGACGT----- 960
DB 1631 AGCCGTGCTATCTTTTGGGGCTGACTCTGTGCGAGCAACACAGAAACAGAGTTGAGC 1690
QY 961 -----CACCTGCT-CCACC-----CCACCC--CCAGTTCTTAAG--GT 995
DB 1691 CTGTTGATGCCCGCTCTGCTGCGAGCGAGACTGACCATCGACAGGTCTTGGGTGA 1750
QY 996 C--TTTTCAGAGC-GTGGAG-----GTGTGGAAGGA-----GTGGCT 1029
DB 1751 CAATATCAAGACGAGAGAGATCTCTGAAGTGAAGATGATGAGAAATTCGACATGACT 1810
QY 1030 -----GCTC-----TCCAAA-----CTATGC----- 1045
DB 1811 CAGGATATGAAGTTTCATCAAAATTTGGTGTCTTTGCAAGATGTGGGTTCAAACA 1870
QY 1046 -----CAAGGC--GGCG-----CAGAGC-----TGCTCTCTGCT 1074
DB 1871 AAGGTGAATCATTTGGACTCATGTGGCGGTGTGTCATAGCGACAGTATCTTC--AT 1928
QY 1075 CTCCTTGAGA-----AAGGTTCTGTGCTGCTGATTTATGAACCTCTATATAGATA 1126
DB 1929 CACCTTGTGATGCTGAGAGAGAAACAG-TACAC-----ATCCATTC-ATCATGTTGTG 1980
QY 1127 TATAGTTTTGTACTTTTTTAC-----AGGAAGTGACTT-TG-----TGTAACAATGC 1175
DB 1981 GTGAGGTTGAGCGGCTGTGTCACCCAGAGAGGCGCCCTGTCAGAGATGCAGCAAC 2040
QY 1176 G---ATGPTATTTAAAC-----TTTTTATAAAGTT 1203
DB 2041 GGCTACGAAATCCAACTACAGTCTCTTTGAGCAGAT 2078

RESULT 4
US-09-548-367D-13
; Sequence 13, Application US/09548367D
; Patent No. 640698
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 29915/6280H
; CURRENT APPLICATION NUMBER: US/09/548,367D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 2088
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-367D-13

Query Match 41.4%; Score 3123.5; DB 4; Length 2089;
Best Local Similarity 37.5%; Pred. No. 8.5e-138;
Matches 780; Conservative 0; Mismatches 403; Indels 895; Gaps 126;

Qy	1	GATCTGGGTGCTGCAGGAAAACCAATTTCTGAACT	52	1101	GGAAAGAGAGCAGCCAAACGAGAGACAGCAGCTGCTGGAGACACATGCGCCAGAGTGGA	1160
Db	21	GCTCTCTGCTGGCGCGCTGGACGGCTCGGCGCTGGAGTACCCACTGATGATGATGCTGG	80	613	AGTCCTG-----GCCG-AGCCAGCCCCCTGGA-----CCCC-----CTGATGCG	649
Qy	53	--TGATTTTAAATCC-----TTGC-----TGCGGAGAG-----GCC--GCCTCT	90	1161	AGCCATGCTCAATGACCGCGCGCTGGCCCTGGAGAACTACATCACCGCTCTCGAGCG	1220
Db	81	CCTGCTGGCTGAACCCAGATTGGCCATGTTCTGTGGCAGACTGAACATGCACATGAATGT	140	650	--CCCCACCTTCCCATCCAGACAGCGGAGCTC-----GC-----TCGG-----G	686
Qy	91	CC-----CGGCT--ATCAGCG--CTTCTCTCATTTCTTTGAATCCGGCG--	128	1221	TGTTCTCTCTCGGCTC--GTCA--CGTGTTCAATATGCTAAAGAAATGATGTCGCGCAG	1276
Db	141	CCAGATGGGAAGTGGGATTCAGATCCATCAGGACCAAAACCTTGCAATTGATACCAAGGA	200	687	AACT-----TGTCATCT--CCACGACAAAGAGAGCTTTGC-----	721
Qy	129	----CTCC--CGGCTCTT--CGGCTGACAGCAG--CCGAGGAGAGCCTGTTTG--CA	174	1277	AACAGAGGACAGACAGACACACCTTAAAGCATTTTCAGCATGTGCGCATGTTGGATCCCA	1336
Db	201	AGGCATCTCGATGATTTGCCAGAGATCTACCTGACTGACATGACATGCAATGTGGTAGA	260	722	-----CACTGACTCGG--CGTGCTCTG--ACACCTCC-----AGAACGCA	758
Qy	175	ATTTAAGCGGGCTGTGAACGCCAGGCGCGGGGGCGGGGCG--AGGCGG--	226	1337	AGAAAGCCGCTCAGATCGGTCCAGGTTATGACACACTCCGTGTGATTTATGACGCA	1396
Db	261	AGCCAACCAACAGTGACCATCCAGAACTGGTGCAAGCGGGCGCCGAAGCAGTGCAGAC	320	759	--GGTGCTCGCGCCGCTTCTGC--CTGGGACCCCGGGAACCT--CTCTCTGCG--GAAGCC	811
Qy	227	-----CCATTTTG-----AATAAAGAGC-----GTGCC--250	250	1397	TGAATCAGTCTCTCTCCCTGCTCT--ACAACGTG--CCTGCAGTGCGCCGAGGAGATTCA	1451
Db	321	CCATCCCACTTTGTGATTCCTTACCCTGCTTAGTGTGTGATTTGTAAGTATGCGCT	380	812	GGACGGCAGGATGGG-----CCCAACT-----TCGCCCT--GCCCA	847
Qy	251	-----TTCCAGGAGG-----CTCTATAAGT-----272	272	1452	GGATGAAGTTGATGAGCTGTCTTCAAGAGAGCAAAACTATTCAGATGACGCTCTTGCCCA	1511
Db	381	TCTCGTTCTGACAGTGGAATCTTACACAGGAGAGATGGATGTTTGGGAACCTCA	440	848	CTTGACTT--CACCAA--ATCCCTTCTCGAGACTAAAC--CTGGTGC-----TCAG	893
Qy	273	-----ACCGCCG-----GGCAGC--GTGCG-----CGCGTTGCAGGT	304	1512	CATGATTAAGTGAACNAGGATCAGTTAC--GGAAACGATGCTCTCATGCCATCTTTGACCG	1570
Db	441	TCTTCACTGGCACACCGTCCCAAGAGAGACATGCAGTGAGAAGAGTACCAACTTGCATGA	500	894	GAGCGAAGG--ACTGTGAA-----CTT-----GT--AGCCTGAA-----G	924
Qy	305	C-AC--TGTAGGGAC-----TTC-----TTTGGTTTT--330	330	1571	AAACGAAACACACCGTGGAGCTCTCCCGTGAATGGAGAGTTTCAGCGCTGACGATCTCC	1630
Db	501	CTACGGCATGTGTGCTCCCTGCGGAATTGCAAGTTCCGAGGGGTAGAGTTTGTGTGTG	560	925	AGCC-----AGAGTAGCTCTG--GCCA--CCAGCTGG-----CGAGCT-----960	960
Qy	331	----CTTTCT-----CTTTGGGCACCT--CTG-----GACTC---357	357	1631	AGCCGTGGCATCTTTTGGGGCTGACTCTGTGCCAGCCAAACACAGAAAACGATTTGAGC	1690
Db	561	CCCCTGGCTGAAGAAAGTGAATGGAATCTGTGATGCGGAGGAGTATGACTCGGA	620	961	-----CACCTGCT--CCCCC-----CCAGCC--CCAAGTTCTAAG--GT	995
Qy	358	-----ACTC--CCCAGCA-----TGAGAGC-----375	375			


```

; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-9

Query Match          41.2%; Score 3105; DB 4; Length 2601;
Best Local Similarity 35.9%; Pred. No. 7,2e-137;
Matches 801; Conservative 0; Mismatches 388; Indels 1040; Gaps 132;

QY      1  GATCTGGGGTGC-----TGC--CA-----GGAAGAGCAAAATTCGAA 37
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     364  GATATGGACTTCACATTAGACTTCTACATCGCTCAAAACGTGGCAAGACCTCGACTAGCC 423

QY      38  GTTAATGGTTTTGA---GTGATTTTT---AAATC-----CTTGTCGCGAGAGG-- 81
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     424  TTCGGAAGTCTTGATTGGGACTTTCCAAAGAATCGACTCACTTACCGTCGAGTAGAC 483

QY      82  -----CCGGCC-----TCTCCCC----- 94
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     484  TACTCGGATAGACTGTGGAACCCGACAGCTTCTCCAAATGAAAGAAATCATTTCTT 543

QY      95  ----GGTATCAGCGC-TTCTCTCAT-TCTTGAATC----- 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     544  CACTTGGCAACACATTAATCTGTTCTTCTGTATCGAGGTGATGGAACGTTTATACT 603

QY     124  ---CGGGCTCCGGGT---CTTGGGGGT---CA---GACCAG-----CCGAGGAA 163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     604  AGTCAAGATTAAACAGTCACTGTCAAGTGTCCAAATGGACCTGAAGCTGTTTCCCAATGGA- 662

QY     164  GCTGTGTTCCAAATTAAGC-GGCTCTGAACGC-CCAGGGCC--GGCGGGGC----- 212
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     663  ---CTCTCAACACTGTAAACTGGAATTAAGACTCGGTACGAGACGAAGAATATCGAC 720

QY     213  -----GGGGCCGAGGGGGCCATTTTGAATAAAGAGGC-----GTGCCCTTC 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     721  TACTATTGGGGAAGACGGGACTGATTTGGAGATAACGGCTGTCAAGTTTGATACCTTC 780

QY     254  CAG---GCAG-----GCTCTA-TAAGTGA-----CCG---CGCG 281
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     781  CAGTTCCCGAGTTTCAGCCAAAGCTGTATTTTGTGAATACAACTAAAGCCGAGACCTCA 840

QY     282  GCGAG-----CGTGGCC--GGTTGTCAGGT-----GGTTCGAAATATGGGCTTCTAC 900
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     841  TCAGGAAATACGTAGCGCTCGCGCTGGAAGTAATATGGTTCGAAATATGGGCTTCTAC 900

QY     306  ACTGT-AGCG--GACTTC-----TTTTGGTTTTCTTT----- 334
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     901  ACTATGAACATCGTCATCCCATCCATCCTGATCGTCACCATATCTTGGGTATCATTTGG 960

QY     335  -----CTC-----TTTCGG-----GCAC----- 347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     961  TTGAATCGAAGCTTCGCGGCTCGAGTTGGTGTGACTACTGTGCTCACAATG 1020

QY     348  ---CTCTGGACTC-ACTCCCA-----GCATGAAGCGCT-----GAGCCCG 385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1021  ACAACTCTGATCACTACAACCAATAATTCGATGCCAAAAGTGTCTTATGTCAAGGGTCTG 1080

QY     386  G-TGCCGGC-----TGCTACGAGGGGTGTGCTGCT-----GTGG-GAACGC--- 427
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1081  GATGTGTTTCTTAATTTTTTGTTCGTAATGGTATTC-GCCTCGTTGCTCGAGTACGCCAT 1139

QY     428  -----AGCTGGCCATCGCCCGGG---GCCGAGGGAAG--GCC-- 461
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1140  AGTATCTACATGAATAAACGACTGTGTCTCCGACGGGAAAAACGAAGAAAGCCGCGGA 1199

QY     462  ---CGCAGC---TGAGAGCGCGCT-----GAGCTTCTGGACCAATGA--- 501
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1200  ACAACAGCAGCAAAACGAGATGCAATGTTTCAACGCGAGGCCGCAAGCCCAATAATAA 1259

QY     502  -----CCACT-----GCTAC--TC-CGCGCTTCGGGAA-CTG-GTACC CGG 537
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB     1260  TTCATAGAAATGACACTTATGTGCGAATAATTCAGCGCTTGCCAAAGACTTGTACAGGC 1319

QY     538  AGTC-----CCGAGGGCAC---TCAGCTTAG--CCA--GGTGAAATTC----- 575
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

1021	ACAAC	TCTGAT	CAC	TACA	CCAA	TAAT	TCGAT	GCCAAA	AGT	GTCT	TAT	GTCA	AGGG	CTG	1080				
386	G	TGCCGGC	-----	TGCT	ACGAG	GGGT	GTCT	GCCT	-----	GT	CG	GAAC	GC	---	427				
1081	GAT	GTG	TTTCT	TAA	TTTT	TGTT	TCG	TAAT	GT	TATC	GCCT	CGT	TGCT	CGAT	1139				
428	-----	AGT	CTG	CCAT	TC	CCCGG	-----	GCC	GAG	GG	GAAG	GC	---	461					
1140	AGT	ATCCT	AT	CAAT	AAAC	GA	CTGT	CTCC	GACG	GGAAAA	ACGA	AAAA	CGCC	CGGA	1199				
462	---	CGC	GAC	---	TGAG	GAG	CGGT	-----	GAG	CTT	CTG	AC	CAC	ATGA	---	501			
1200	ACA	ACAG	CGAA	CG	AGAT	GCC	CAAT	GT	CAAC	GGAG	CGCC	GAG	CGCC	CAATA	TAA	1259			
502	-----	CCACT	-----	GCT	AC	---	TC	CGC	CTG	GGGA	CTG	GT	AC	CCCG	537				
1260	TT	CAT	ACGA	AT	GAC	ACT	TAT	GT	CG	CAAA	AT	TC	AG	CCCT	TAT	GT	AC	AGC	1319
538	AGTC	-----	CCG	AGG	GCAC	---	TC	AGT	TTAG	---	CCA	---	GGT	GAAA	TCC	---	575		

Db 1320 TGACTTGTACTTTGGCGG---GACACAAATTCCTCTATGAATCCATTTGATGAGATCCGAGA 1376
QY 576 -----TACAGCGGT---CATCGA---CTACATTTCTCGACCTG---CAGGTAGTGC 617
Db 1377 AAATTGTGATGCGCGAGATCCAAATGATGATCAACATCCAGCTTGTGTACAGAGCGGC 1436
QY 618 -----TGGCGG-----AGC-CAGCC-----CTTGGAC-----CCC 641
Db 1437 ACATACGCTATGCGCGGTCCATTTGCGCGCGCGGAAAGAGCTTCCAAAGACATGCTGCCA 1496
QY 642 CTGATGGCCCGC-----ACC-----TTCCCA----- 662
Db 1497 AGATGGAGCGCTGCAAAATCGATAAGCTTAGCGGATACGGTTTCCCATTTGCTTCTC 1556
QY 663 -----TCGAGACAGCC-----GAGTCTCGT---CCGGAACCTTGTG---A 695
Db 1557 TATCTTCAATATAGTCTACTGTTGTATATGAAATATCTAAGCTTAACTCGTGGCAAA 1616
QY 696 TCTCC-----AACGACAAAGG-AGCTTTTCCCACTGA-----C-TGGGCG----- 734
Db 1617 GATCCAGGAGACGACAGTGGCAGCAT-CCACTGATGGTATTCGACGGCGGAAATC 1675
QY 735 -----GTGTCC-----TGACACC--TCCAGAACGCGAG-- 759
Db 1676 GAGTCAAAATGCTGTACGTCGAAGGAGCGGAATTTGTCGACAGCGGTCAAGGCGGACGG 1735
QY 760 -----GTGCTGGCG-----CCC-----GTTCTGCTG-----GGACC----- 787
Db 1736 AACATCGCACTCTCAGTTATAATCTACTAATCTGCGAAGACGACACTTGCAGC 1795
QY 788 -----CGGGAAC---CTCTC---CTGC-----CGGAAG----- 809
Db 1796 ACTTCATCGGACCTACTCTGCTACGGGTAGTTTCATATTTGATCGACAGCGG 1855
QY 810 -----CGGACCG-----CAGGGAT----- 824
Db 1856 TTCTACTTTCTCAAAATTTTCCCTGCGAGCTGCTGCTAGTTTATCATGATCTCA 1915
QY 825 -----GGGCCC---CAACTTCG---CCC-----TGCCC 846
Db 1916 TTCTGGATCAATCGTGACTCGCGGCTTTCGGAACCTTAATCGGTACGATGACGGTGCTC 1975
QY 847 ACT-----TGAC-----TTACCA----- 860
Db 1976 ACTGAGACTCTTTATGACCGGAACCAATCGAGCTCTTCCACCAGTTGCTATGTAAA 2035
QY 861 -----AATCC-----CTTC-----CTGG-----AGACTAA 880
Db 2036 GCGGTGATGATTCCTCGGTTTCTGCTATCTTCTGTTATCTGCGGTTGATCGAGTAC 2095
QY 881 ACTTG-----TGCTCAGGAGCGAAGGACTGTGAATTT---GTAGCTTGAAG-AGCCA 929
Db 2096 GCTGTGTTGCTACTCATAAAGAGAAACGAGGATCGTCGAGAGAGAGAGAGAGC 2155
QY 930 GAGCT-----AGCTCG---GCCACCAG-----CTGGGCGAGCTCACCT----- 966
Db 2156 GAGCAATAACCTGCTCCCGCAGACCTGATATTTCTCAGCAGCTCGCCCTTGGCGAATGC 2215
QY 967 -----GCTCCACC-----CCACCCCGCAAGTTCTAAGGTCTTTTTCAGA 1004
Db 2216 ACATCAACGCGGCTCCAACTCGATATCGCGCTCATCAAGCAGTCCGATCGATTCTGT 2275
QY 1005 G-----CGTGGAGGTGGAGAGGTGSC-----TGCTC----- 1033
Db 2276 GTCAGTCAAGTCAATTGACAT-CGTCAGCGGTGCGCGGTTCTCTTGTGTTTATCTT 2334
QY 1034 ---TCCAAACTATGCCAAGCGG---CGGCAGAGCTGCTTCTGCT-CC-----TTGG 1082
Db 2335 GTTCAACACTCTCTCTGCTGATCTTACTGTCTACAAATCCAAAGCGCTGCGCTGATATTAG 2394
QY 1083 AGAA---AGGTT---CTGTTGC-----CTTGA---TTTATGA---CTCTAT----- 1117

Db 2395 TGAACAGAGGCTGACCGTTGCGATGCTCCAGACCTTCATTATCTCAATCCAATTCCT 2454
QY 1118 -----AATAGATATAGGTT---TTGTAC-----CTTTTATAC-----AGGA 1153
Db 2455 CATCAITTTCCATTTGCAATATCTTTTCTTTCGACAGAGCTTTTTCGTTTTTTT 2514
QY 1154 AGGTGACTTTCTGTAAC-----AATGC-----GATG-----TATATTAAC 1189
Db 2515 TATTGATTATTTTACGGATTTTATAGTAATACAGATGCTCATTCTCAATAAAT 2574
QY 1190 TTTTATAA 1198
Db 2575 TTTTATAA 2583

RESULT 7
US-09-436-063C-9
; Sequence 9, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436, 063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-9

Query Match 41.2%; Score 3105; DB 4; Length 2601;
Best Local Similarity 35.9%; Pred. No. 7, 2e-137;
Matches 801; Conservative 0; Mismatches 388; Indels 1040; Gaps 132;

QY 1 GATCTGGGGTGC-----TGC-CA-----GGAAAAGCAAAATTCGGAA 37
Db 364 GATATGACTTCACATTTAGACTTCTACATGCTGTCAAACGTCGCAAGACCTCGACTAGC 423
QY 38 GTTAAUGTTTGA---GTGATTTT---AAATC-----CTTGTGCGGAGG--- 81
Db 424 TTGGAAGTCTTGTATTTGGGACTTTCCAAAGAAATCGACTCACTTACCCTCGGAGTAGAC 483
QY 82 -----CCGCGC-----TCTCCCC----- 94
Db 484 TACCTGATAGACTGTGGAACCCGACAGCTTCTCCCAATGAAAAGAAATCATTTCTTC 543
QY 95 -----GCTATCAGCGC---TTCCTCATT---CTTTGAATC----- 123
Db 544 CACTTGGCAACACACATTAATCTGTTCTTCTGATGAGGGTGTGGAACGGTTTATCT 603
QY 124 ---CGCGGCTCCCGGT---CTTCGGCGT---CA---GACCAG-----CCGAGAGAA 163
Db 604 AGTCAAGATTAACAGTCACTGCAAGCTGTCCAAATGGAAGCTGGAAGCTGTTCCTCAATGA 662
QY 164 GCCTGTTTGAATTTAAGC---GGCTGTGAACGC---CCAGGGCC---GGCGGGGC----- 212
Db 563 ---CTCTCAACACTGTATACTGGAATTTGAAGCTAGGGTACGACGCAAGATATCGAC 720
QY 213 -----GGGCGCGAGGGCGCCATTTTGAATAAAGAGC-----GTGCTTTC 253
Db 721 TACTATTGGGGGAAGAAGCGGACTGATTTGGAGATAACGGCTGTCAAGTTTGTATACCTTC 780
QY 254 CAG-----GCAG-----GCTCTA---TAAGTGA-----CCG---CCGG 281
Db 781 CAGTTGCCGCGAGTTTCAGCCACGCTGTATTTTGTGATACAACTAAAGCCGAGACTCA 840

Db	1916	TTCTGGATCAATCGTGA	CTCGGCGCCTTCGGAACCTTAATCGGTACGATGACGGTGCTC	1975	
Qy	847	ACT	TTGAC	TTGACCA	860
Db	1976	ACTGAGACTCATCTTATGACCGGAACCAATCGAAGTCTTCCACGATTCGCTATGTAAAA	2035		
Qy	861	-----AATCC-----	CTTC-----	CTGG-----	AGACTAA 880
Db	2036	GCGCTGATGATTA	CTCTCGGTTCTGCTATCTTCTGGTTTAACTTGGCGTTGATCGAGTAC	2095	
Qy	881	ACTGG-----	TGCTCAGGACGGAAGACTGTGA	CTT-----	GTAGCCTGAAG-AGCCA 929
Db	2096	GCCTGTGTTGCC	TACTCAAAAAAGAAACGAGGATCGTCGGAAGAAGAGAAGAAGACG	2155	
Qy	930	GAGCT-----	AGCTCTG-GCCACGAG-----	CTGGCGGAGCTCACCCCT-----	966
Db	2156	GAGCATAAACCTGCTCCGCGACACCTGATATTC	TTCTTACGAGCTCCGCTTGC	CGAATGC 2215	
Qy	967	-----GCTCCGACC-----	CCACCCCAAGTTCTAGGTCCTTTTCAGA	1004	
Db	2216	ACATGCAACGGCGCTCCAACTCGATCATCGCGCTCATCAAGACGTCGAA	TGATCTCTGT 2275		
Qy	1005	G-----	CGTGAGGTGTGGAAGAGTGGC-----	TGCTC-----	1033
Db	2276	GTCAATCACAGTCACATTGACAT	CGTCAGCGGTGCGCGTTCTCTTGTGTTCACTTT 2334		
Qy	1034	-----TCCAAATATGCCAAGGGG-----	CGGACAGCTGGTCTTCTGCTCT-CC-----	TTGG 1082	
Db	2335	GTTCACACACTCTCTTCTGGCTGATCTACTGTACAAATCCAAAGCTGTGCGGTATATTAG	2394		
Qy	1083	AGAA-----	AGGTT--CTGTTC-----	CCGTA--TTTATGAA-CTCTAT-----	1117
Db	2395	TGAACACGAGGGTGACGGTTGGATGCTCCAGACCTTCATTAATCTCAATCCAACTTCCT	2454		
Qy	1118	-----AATAGAGTATATAGTT--	TTGTAC-----	CTTTTATAC-----	AGGA 1153
Db	2455	CATCATTTTCCATTTCCGAATATCTCTTTTCTTGACAGAAGCCCTTTTTCGTTTTTTT	2514		
Qy	1154	AGGTGACTTTCTGTAAAC-----	AAATGC--GATG-----	TATATATAAC 1189	
Db	2515	TATTGATTTATTTTACGGATTTTATAGATAATGACAGATGCTCATCTGCTCAATAAAT	2574		
Qy	1190	TTTTTTATAA	1198		
Db	2575	TTATTTTAA	2583		

RESULT 8

RESOLUTION 8
US-09-627-650B-7
; Sequence 7, Application US/09627650B
; Patent No. 6406872

; Patient NO. 640667Z
; GENERAL INFORMATION:
; GENERAL INFORMATION:

; GENERAL INFORMATION:
 ; APPLICANT: Bamber, Bruce
 ; APPLICANT: Jorgensen, Erik
 ; TITLE OF INVENTION: Nematoe Neuromuscular Junction GABA Receptors and
 ; TITLE OF INVENTION: Methods Related Thereto

FILE REFERENCE: 21101-0009U3

FILE REFERENCE: Z1101.000903
CURRENT APPLICATION NUMBER: US/09/627.650B

; CURRENT APPLICATION NUMBER: 03/03/03
 : CURRENT FILING DATE: 2000-07-28

;; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063

;; PRIOR APPLICATION NUMBER: 037
: PRIOR FILING DATE: 1999-11-08

;; PRIOR FILING DATE: 1999-11-08
: PRIOR APPLICATION NUMBER: 60/107,727

; PRIOR APPLICATION NUMBER: 607
 ; PRIOR FILING DATE: 1998-11-09

; PRIOR FILING DATE: 1998-11-
; NUMBER OF SEQ ID NOS: 50; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.1; SOFTWARE: E
: GEO ID NO 7

```

; SEQ ID NO 7
      LENGTH: 2508

```

```
; LENGTH: 25
; TYPE: DDT
```

TYPE: PRT
OPCARISM: Caenorhabditis elegans

; ORGANISM: Ca
NC 08-637-6E0B-7

Give Match

41.1%: score 3097.5; DB 4; Length 2508;

282 GGAG-----CGTGGC---GGTTGCGAGT-----C 305
 841 TCAGGAAATACGTACGCGCTGGCGCTGGAAGTAAATATGGTTCGAAATATGGCTCTTAC 900
 306 ACTGT-AGCG--GACTTC-----TTTGTGGTTTCTTT-----334
 901 ACTATGAACATGTCATCCCATCCATCCTGTGCGTCACCATATCTTGGGTATCATTTTGG 960
 335 -----CTC-----TTTGGG-----GCAC-----347
 961 TTGAATCGAGAACTTCGCGCGCTCGAGTTGGATTGGGTGTGACTACTGTCTCACAATG 1020
 348 -----CTCTGACTC-ACTCCCCA-----GCATGAAGCGCT-----GAGCCCG 385
 1021 ACAACTCTGATCACTACAAACCAATAATCGATCCAAAAGTGTCTTATGTCAAGGTCTG 1080
 386 G-TGCGGGC-----TGCTACGAGGCGGTGTGTGCT-----GTGG-GAAAGC-----427
 1081 GATGTGTTCTTAATTTTGTGTTTCGTAATGGAATTC-GCCTCGTGTCTGAGTACGCCAT 1139
 428 -----AGTGTGGCCATCGCCCGGG-----GCCGAGGAGAGG--GCC--461
 1140 AGTATCCTACATGAATAAAGCACTGTGTCTCCGACGGGAAAAACGAAGAAAGCGCCGA 1199
 462 -----CGGAGC-----TCAGGAGCGCT-----GAGCTGCTCGAGACATGAA---501
 1200 ACAACAGCGGAAAACGATGCCAATGTTCAACGCGAGGCCGAAAGGCCCAATATAA 1259
 502 -----CCACT-----GCTAC--TC-CGCGCTGCGGAA--CTG-GTACCCGG 537
 1260 TTCATACGAAATGACACTTATGTGCGCAAAATTCGACGCTTCGCAAAAGCTATGTACAGC 1319
 538 AGTC-----CCGAGAGGCAC--TCAGCTTAG--CCA--GGTGGAAATCC-----575
 1320 TGACTTGACTTTGCGG---GACACAATTCCTCTATGAATCCATTGATGAGATCCAGA 1376
 576 -----TACAGCGCGT---CATCGA--CTACATTTCTGCAGCTG---CAGGTAGTCC 617
 1377 AATTGTGATTTGCGGACGATTCCAATGATGACACA'CCACGCTTCTCACAGACGCGC 1436
 618 -----TGGCGG-----AGC-CAGCC-----CCTGGAC-----CCC 641
 1437 ACATACGCTATGGCGGCTCCATTCGGCGGCGCGAAAGGCTTCCAAAGACATGTGCCA 1496
 642 CTGATGGCCCC-----AGC-----TTGCCA-----662
 1497 ACGATGAGCGCTGCAAAATCGATAAGCTTAGCCGATAGCGTTCCCAITGCTTCTC 1556
 663 -----TCAGACAGCC-----GAGCTCGCT---CCGGAATCTGTC-----A 695
 1557 TATCTTCAATATAGTCTACTGGTTGTATPATAAATATCTAAGCTTAAACTGTCGCGCAA 1616
 696 TCTCC-----AACGACAAAAG-AGCTTTTCCCACTGA-----C-TGGGCC-----734
 1617 GATCCAGAGAACGACAGTGGCAGCAT--CCA'CTGATGCGTATTCGACGGCGGAAATC 1675
 735 -----GTGPTCC-----TGACACC--TCCAGAACGAG--759
 1676 GAGTACAAATGGTGTACGTTCGAAGGAGCGCAATTTGTCGACAGCGGTCAAGCGCGACGG 1735
 760 -----GTGCTGGCG-----CCC--GTTCTGCCTG-----GGACCC-----787
 1736 AACATCGAACTCTGAGTTTAAATCTACTAAATCTGCCAAAACGACACTTGCAGC 1795
 788 -----CGGGAAC--CTCTC--CTGC-----CGGAAG-----809
 1796 ACTTCATCGGGACCTACTCTGCTACGGGTTAGTTTCATATTTGATTCGCGACAGGGC 1855
 810 -----TCCGAGCGG-----CAGGGAT-----824
 1856 TTCTACTTTC'CAAAATATTTTCCCTCGCAGCCTCGTGTAGTTTATCATGTGATCTCA 1915
 825 -----GGGCCC--CAACTTGG-----CCC-----TGCCC 846

Best Local Similarity 36.5%; Pred. No. 1.6e-136; Matches 779; Conservative 0; Mismatches 409; Indels 949; Gaps 120;		
QY	1	GATCTGGGTGC-----TGC--CA-----GGAAGGCAAAATTTCTGAA 37
Db	364	GATATGAGCTTCACTATGACTTCTACACTGCTCAACGTCGCAAGACCTCGACTAGCC 423
QY	38	GTAAATGGTTTGA---GTGATTTT-----AAATC-----CTTGTGGCGGAGG-- 81
Db	424	TTGGGAAGTCTTGAATTTGGGACTTTTCCAAAGAAATCGACTCATTACGTCGGAGTAGAC 483
QY	82	-----CCCGCC-----TCTCCCC----- 94
Db	484	TACCTGGATAGACTGTGGAACCCGACACGTTCTTCCCAATGAAAGAAATCATTTCTTC 543
QY	95	-----GGTATCAGGC--TTCCTCAT--CTTTGAATC----- 123
Db	544	CACCTGGACACACATCACTAACTCGTTCCTTGTATCGAGGGTGATGAAACGTTTATACT 603
QY	124	---CGGGCTCGCGGT--CTTCGGGT---CA---GACCAG-------CCGAGGAA 163
Db	604	AGTCAAGATTAACAGTCACTCAACGCTGTCATGAGACCTGAAGCTGTTCCTCAATGGA- 662
QY	164	GCCTGTTTCAATTAAGC--GGCTGTGAAGC--CCAGGCC--GGGGGGG-- 212
Db	663	--CTCTCAACACTGTAACTGGAATTTGAAATGAACTACGGGTAGAGACGAAGATATCGAC 720
QY	213	-----GGGCGGAGGCGGCCAATTTGAAATAAGAGGC-----GTGCTTC 253
Db	721	TACTATTGGGGGAAGAGCGGACTGATTTGGAGATAACGGGTGTCAAGTTTGTATACCTTC 780
QY	254	CAG-----GCAG-----GCTCTA--TAAGTA-----CCG-----CCGG 281
Db	781	CAGTTCCGCGGTTTCAGCCACAGCTGTATTTTGTGAAATACAACTAAGCCGACCTCA 840
QY	282	GCAG-----CGTGGC--GGTTGCGAGT----- 305
Db	841	TCAGGAAATACGTACGCTGGCGTGGAGTAATATTGGTTTCAAAATATGGGCTTCTAC 900
QY	306	ACTGT--AGCG--GACTTC-----TTTTGTTTTCTTT----- 334
Db	901	ACTATGACATCGTCACTCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 960
QY	335	-----CTC-----TTTGGG-----GCAC----- 347
Db	961	TTGAATCGAAGACTTCGCGCTCGAGTTGGATTGGGTGTGACTACTGTCTCACAATG 1020
QY	348	-----CTCTGGACTC--ACTCCCA-----GCATGAAGCGCT-----CAGCCCG 385
Db	1021	ACAACTGTGATCACTACACCAATAATTCGATGCCAAGGTGTCTTATGTCAAGGCTCG 1080
QY	386	G--TGCGGGC-----TGCTACGAGGCGGTGTGCTGCTT-----GTGCG--GAAGC-- 427
Db	1081	GATGTGTTCTTAAATTTTGTTCGTAATGTTTTC--GCCTCGTTCGAGTAGCCAT 1139
QY	428	-----AGTCTGGCCATCGCCCGG--GCCGAGGGAAG--GCC-- 461
Db	1140	AGTATCTTACATGAATAACGACTGGTCTCCGACGGGAAACGAAGAAAGCCGCGA 1199
QY	462	-----CGGCAGC-----TGAGAGCGGT-----GAGCTTGTGAGAGC-----A 497
Db	1200	ACAACAGCAGCAAGATGCAATGTTCAACGCGAGCCGGAAGCGCCCAATAATAA 1259
QY	498	TGAACCACTGCT---ACTCCC-----GCCTGCGGGA-----ACTGGTAC----- 533
Db	1260	TAATCCATGTATGAGATCCAGAAATTTGTGATGTCGCGGAGATTCCAAATGATCAACA 1319
QY	534	CCG-----GAGTCCGAGGAGCACTCA---GCTTAGCCAGGTGGAATCTTACAGCGGT 585
Db	1320	TCCAGGTCTGTACAGAGCGGCGACATAGCTATGGCGG-----CTCCATTGCGCGG 1374
QY	586	C-----ATGACT-----ACATTCT-----CTGCA-----GGTAGTCTCT 618

Db	1375	CGAAAAAAGGTTCCAAAGACATGCTGCCAAGCATGAGTGGACGGCTGCCAAAAATCGATAAGCTT 1434
QY	619	GGCGAGCCAG---CCC-----CTGG----- 636
Db	1435	AGCGATACGTTTCCCAATGCTCTTCTCTATCTTCAATATAGTCTACTGTTGTATATG 1494
QY	637	-----ACCC-----CCTGATGCCCCCA---CTTCCATC 664
Db	1495	AAATATCTAAGCTTAAATCTGTCGGAACAAGATCAGGAGAACACAAGTGGCAGAGATC 1554
QY	665	CA-----GACAGCGGAGCTCG-----CTCGGAAC----- 689
Db	1555	CACATGATCGTATTCGACGGCCGAAATCGAGTAGTCAAAATGGTGTACGTCGAAGGAGCGAA 1614
QY	690	---TT-----GTCACTCTCCACGACAAAG--GAGCTTTTGC-----CACT--G 726
Db	1615	TTGTTTCAAGCGGTGAAGCCGACGCACTGCAACTGTGAGTTATAATTTCACTAA 1674
QY	727	ACTGCGCGTGTCTGACACCT--CCAGAAC-----GACGG-- 760
Db	1675	AATCTGCCAAAAAGCGACACTTGCACGCACTTCATCGGGGACCTACTCTGCTACGGGT 1734
QY	761	-----TGCTGGC-----CCGTTCTG 777
Db	1735	TAGTTTCATATTTGATCGGACAGCGCTTCTACTTTCTTCAAAATATTTTCCCTCCAG 1794
QY	778	CC-----TGGGACCC---CGGGAAC-----CT---CTCCTGCCGG 806
Db	1795	CCTCGTGTAGTTTATCATGATCTCATTTCTGATCAATCGTGAATCGGCGCTTCGCG 1854
QY	807	AAGCG-----GACGCGAGG-----ATGGGCCCAACTTC--GCCGTG--CCCACTG 851
Db	1855	AACCTAATCGGTACGATGACGCTGCTCACTGAGACTCATCTTATGACGGAACCAATCG 1914
QY	852	AC---TTCAACA-----AATCC-----CT 867
Db	1915	AGTCTTCCACAGTTGCTATGTAAAGCGGTGTGATGTATCTCTCGGTTTCTGTATCT 1974
QY	868	TC-----CTGG-----AGACTAAACCTCG-----TGCTCAGGAGCGAGACTG 906
Db	1975	TCTGTTATATCTGGGTTGATCGAGTACGCTGTGTGCTACTCTCAAAAAGAGAACGA 2034
QY	907	TGAAGTT---GTAGCCTGAAG--AGCCAGAGCT---AGCTCTG--GCCACCAG--- 949
Db	2035	GGATCGTCGGAAGAGAGAGAACGAGACGAGCATAAACCTGCTCCCGCCGACCTGATAT 2094
QY	950	---CTGGCGACCTCAACCT---GCTCCACC-----C 976
Db	2095	TCTTCAAGCTTCGCGCTTGGCGAATGCACATGCAACGCGGCTCCAACCTCGATATCGC 2154
QY	977	CACCCCAAGTTCTAAGTCTTTTCAGAG-----CGTGGAGTGTGGAAGA 1023
Db	2155	CGTATCAAGCAGTCGAATCGATTCTGTCTGACGTCACAGTCACATTGACAT--CGTCAGCC 2213
QY	1024	GTGGC-----TGCTC-----TCCAAACTATGCCAAGGCGG---CGGCAGA 1060
Db	2214	GTGCGCGTTTCTTCTTGTTCATCTTGTTCACACTCTCTTCTGCTGATTTCTACTGT 2273
QY	1061	GCTGCTTCTGCTCT--CC-----TTGGAGAA---AGGTT--CTGTTGC-----CCTG 1102
Db	2274	ACAAATCAAGCTGTGCGGTATATAGTGAACAGAGGTGACCGTTGGGATGCTCCAG 2333
QY	1103	A--TTTATGAA--CTCTAT-----AATAGAGTATATAGGTT--T 1135
Db	2334	ACCTTCAATTAATCTCAATCCAACTTCCTCATCTTTTCCATTTTCCGAATATCTCTTTTCT 2393
QY	1136	TGTAC-----CTTTTTTAC-----AGGAAGTGAATTTCTGTAAC-----AAT 1173
Db	2394	TGCACAGAGCTTTTTCGTTTTTTTTTTTATTTATTTATTTTTCGGAATTTTAGATAAT 2453
QY	1174	GC---GATG-----TATATTAACCTTTTATAA 1198
Db	2454	GCACAGATGCTCATGCTCAATAAATTTATTTTAA 2490

RESULT 9

```

US-09-436-063C-7
; Sequence 7, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-7

```

Query Match 41.1%; Score 3097.5; DB 4; Length 2508;
 Best Local Similarity 36.5%; Pred. No. 1.6e-136; Indels 949; Gaps 120;
 Matches 779; Conservative 0; Mismatches 409;

QY	1	GATCTGGGGTGC-----TGC--CA-----GGAAAAAGCAAAATTCGGAA	37
DB	364	GATATGGACTTCACATTAGACTTCTACATGGCTCAAACGTGGCAAGACCCTCGACTAGCC	423
QY	38	GTTAATGGTTTGA---GTGATTTT-----AAATC-----CTTGCTGGCGAGAGG--	81
DB	424	TTCGGAAGTCTTGATTTGGGACTTTCCAAAGAAATCGACTTACCTACCGTCGGAGTAGAC	483
QY	82	-----CCGCGC-----TCTCCCC-----	94
DB	484	TACCTGGATAGACTGTGAAACCCGACACGTTCTTCCCAATGAAAGAAATCATTTCTC	543
QY	95	-----GGTATCAGGC---TTCCTCATTT---CTTTGAATC-----	123
DB	544	CACATTGGCAACACACATAACTGTTCTTCTGATCGAGGGTGTATGAACGGTTTATPACT	603
QY	124	---CGCGCTCCGGGT---CTTCGGCGT---CA---GACCAG-----CCGGAGGAA	163
DB	604	AGTAAAGATTAAACAGTCACTGCAACGTCCTCAATGACACCTGAAGCTGTTCCTCCATGGA-	662
QY	164	GCCTGTTTGAATTAAAGC---GGGCTGTGAACGC---CCAGGGCC---GGCGGGGGC-----	212
DB	663	---CTCTCAACACTGTAAACTGGAATTTGAAAGTACGGGTACGAGACGAAAGATATCGAC	720
QY	213	-----GGGCGCGAGCGGGCCATTTTGAATAAAGGC-----GTGCCTTC	253
DB	721	TACTATTGGGGGAAGACGCACTGATTTGGAGATAACGGCTGTCAAGTTTGATACCTTC	780
QY	254	CAG-----GCAG-----GCTCTA-TAAGTGA-----CCG---CCGCG	281
DB	781	CAGTTCCGCGAGTTTCACCAACGCTGTATTTTGTGAATACAACTAAGCCGAGACCTCA	840
QY	282	GGAG-----CGTGGC---CGGTTCAGGT-----C	305
DB	841	TCAGGAAATACGTACGCTCGCGCTGGAAGTAAATATTTGTTTCGAAATATGGGCTTCTAC	900
QY	306	ACTGT-AGCG-GACTTC-----TTTGGTTTCTTT-----	334
DB	901	ACTATGAACATCGTCATCCCATCCATCTGTATCTGATCTGCACCATACTCTGGGTATCATTTGG	960
QY	335	-----CTC-----TTTGGG-----GCAC-----	347
DB	961	TTGAATCGAAGACTTCGCGGCTCGAGTTGGATGGTGTGACTACTGTGCTCACAATG	1020
QY	348	-----CTCTGGACTC-ACTCCCA-----GCATGAAGGGCT-----GAGCCCG	385

DB	1021	ACAACTCTGTACTACTACAAACCAATAAATTCGATGCCAAAAGTGTCTTATGTCAAGGGTCTG	1080
QY	386	G-TGCGCGGC-----TGCTACGAGCGGTGTGCTGCCT-----GTGCG-GAAGCG---	427
DB	1081	GATGTGTTTCTTAATTTTGTTCCTAATGATTC-GCCTCGTGTCTCGATAGCCCAT	1139
QY	428	-----AGTCTGGCCATCGCCCGG---GCCGAGGGAAG--GCC--	461
DB	1140	AGTATCTTACATGAATAAAACGACTGGTCTCGACGGGAAAAACGAGAAAGCCGCGA	1199
QY	462	-----CGGAGC---TGAGAGCGCGCT-----GAGCTTGTCTGAGCGAC-----A	497
DB	1200	ACAACAGCAGCGGAAACGAGATGCCAATGTTCAACGAGAGCCGCAAGGCGCGCAATAATAA	1259
QY	498	TGAACCACTGCT---ACTCCC-----GCCTGCGGGA-----ACTGGTAC-----	533
DB	1260	TAATCCATTGATGGAGATCCCGAATAATTGATTGCCGAGCATTCCTCAATGATGCAACA	1319
QY	534	-CCG-----GAGTCCGAGAGGCACTCA---GCTTAGCCAGGTGGAATCTTACAGCGGT	585
DB	1320	TCCACGTCTTGTACAGAGCGCGCACATACGCTATGGCCGG-----TCCCATTCGCGGG	1374
QY	586	C-----ATCGACT-----ACATTC-----CGA-----CCTGCA-----GGTAGTCT	618
DB	1375	CCGAAAGAGCTTCCAAAGACATGTCGCAACGATGGAGCCCTGCAAAATCGATAAGTT	1434
QY	619	GCCGAGCGCAG---CCC-----ACCC-----CTGTATGATAGTCTACTGTGTATATG	1494
DB	1435	AGCCGATACGGTTTCCCATTTCTCTATCTTCAATATAGTCTACTGTGTGTATATG	1544
QY	637	-----GACAGCGAGCTCG-----CTCGGGAAC-----	689
DB	1495	AAATATCTAAGCTTAAACTCTGTCGGACAAAGATCCAGGAGAGCAGAGTGGCAGCAGATC	1554
QY	665	CA-----GACAGCGAGCTCG-----CTCGGGAAC-----	689
DB	1555	CACGTATGCGTATTTCGACGGCGGAAATCGAGTACAAATGGTGTACGTGCAAGGAGCGGAA	1614
QY	690	---TT-----GTCATCTCCACGACAAAG-GAGCTTTTGC-----CACT-G	726
DB	1615	TTGTTCGACAGCGGTCAAGCGCGACGCAACATCGACTGTGCGATTATAAATTCACATAA	1674
QY	727	ACTCGGCGGTGTCCTGACACCT-CCAGAAC-----GCAGG-760	760
DB	1675	AATCTGCCAAAAACGACACTTGCAGCACCTTCATCGGGGACCTACTCTCGTCTACGGGT	1734
QY	761	-----TGCTGGG-----CCGTTCTG777	777
DB	1735	TAGTTTCATATTGTATCGCGACAGCGGTTCTTACTTTCTTCAAAATATTTTCCCTGCGCAG	1794
QY	778	CC-----TGGGACC---CGGGAAC-----CT---CTCCTGCGCG806	806
DB	1795	CCTCGTGTAGTTTATCATGATCTCATTTCTGGATCAATCGTACTCGCGCTTTCGCG	1854
QY	807	AAGCG-----GACGCGAGG-----ATGGGCCCAACTTC-GCCGTG-CCCACTTG	851
DB	1855	AACCCCTAATCGGTAGATGACGGTGTCTACTGAGACTCATCTTATGACCGGAACCAATCG	1914
QY	852	AC---TTCACCA-----AATCC-----CT867	867
DB	1915	AGTCTTCCACAGTTGCTATGTAAGCCGTTGATGTATCTCTCGGTTTCTGTATCT	1974
QY	868	TC-----CTGG-----AGACTAAACCTGG-----TGCTCAGAGGAGGAGCTG	906
DB	1975	TCTGGTTACTTGGCGTTTATCGAGTACGCTGTGTGCTACTCAAAAAGAGAGAGCA	2034
QY	907	TGAACCT---CTAGCCTGAAG-AGCCAGAGCT---AGCTCTG--GCCACAG---949	949
DB	2035	GGATGTTCGGAG	2094
QY	950	-CTGGGCGACGTCAACCT-----GCTCCACC-----C976	976

2095	Db	TTCTCAGCAGCTCGCGCTTGGCGAATGCATGCAACGCGGTCCAACTCGATCATCGC	2154
977	Qy	CACCCCCAAGTTCTAAGGTCTTTTCAGAG-----CGTGGAGGTGTGGAAGGA	1023
2155	Db	CGTATCATCAAGCAGTCGAATCGATTCTGTGCAGTCACAGTCACATTGCAT-CGTCA	2213
1024	Qy	GTGGC-----TGCTC-----TCCAAACTATGCCAAGCGG---CGGCAGA	1060
2214	Db	GTGCGCGGTTTCCTCTCTGTGTTTTCATCTTGTTCACACATCTCTCTCGGTGA	2273
1061	Qy	GTGTGCTCTGGTCT-CC-----TTGGAGAA---AGGTT--CGTTGC-----CGTG	1102
2274	Db	ACAAATCCAAAGCGTCTGCCGTATATTAGTGAACACGAGGCGTACCGTTGCGATGCT	2333
1103	Qy	A--TTTATGAA-CTCTAT-----AATAGAGTATATAGGTT--T	1135
2334	Db	ACCTTCATTATCTCAATCCGAATCTCTCATCATTTTCCATTTTCGAATATCTTTTCT	2393
1136	Qy	TGTAC-----CTTTTTCAC-----AGGAAGTCACTTCTGTAAAC-----AAT	1173
2394	Db	TGCACAGAAGCGCTTTTTCGTTTTTTTTTTTATGATTTATTTTACGGATTTTATAGAT	2453
1174	Qy	GC---GATG-----TATATTAAACTTTTTTATAA	1198
2454	Db	GCACAGATGCGCTCATTTGCTCAAAATAAATTTATTTTAA	2490

RESULT 10

```

RESUL 10
US-09-627-650B-3
; Sequence 3, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-06
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-3

```

Query Match	41.1%	Score 3097.5;	DB 4;	Length 2544;
Best Local Similarity	36.1%;	Pred. No. 1.6e-136;		
Matches 784;	Conservative 0;	Mismatches 404;	Indels 985;	Gaps 125;
Qy	1	GATCTGGGGTGC-----TGC--CA-----GGAAAAGCAAAATTCGAA	37	
Db	364	GATATGGACTTCACATTAGACTTTCATGTCGTCAAACGTGGCAAGACCTCGACTAGCC	423	
Qy	38	GTTAATGTTTGA--GTGATTTTT-----AAATC-----CTTGCTGGCGGAGAG--	81	
Db	424	TTCGGAAGTCTTGATTGGGACTTTCAAAGAAATCGACTCATCTTACCGTCGGAGTAGAC	483	
Qy	82	-----CCGCCC-----TCTCCCC-----	94	
Db	484	TACCTGGATAGACTGTGGAAACCGCAGCTTCTCCAAATGAAAGAAATCATCTTC	543	
Qy	95	-----GGTATCAGCGC-TTCCCTCATT-CTTTGAATC-----	123	
Db	544	CACCTTGGCCAACCACACATAACTCGTTCCCTTCGTATCGAGGGGTGATGGAAACGTTTATACT	603	
Qy	124	----CGCGGCTCGGGGT--CTTGGCGT--CA--GACGAG-----CCGAGGAA	163	

D	b	604	AGTCAAGATTAA	CAGTCACTG	CAACGCTG	CCAAATGG	GCACCTG	GAAGCTGT	TCCCAATGGA	662	
Q	y	164	GCCTGTTTGG	CAATTTAAGC	GGGCTGTG	AAGCGC	CCAGGGGC	GGGGGGGC	-----	212	
D	b	663	CTCTCAAC	CACTGTA	AACTGGA	AAATTGA	AGAGCTAC	GGGTAC	GAGACGAA	720	
Q	y	213	-----	GGGGCCGAGG	CGGGCCATTT	TGAATAAG	AGGCGC	-----	GTGCGCTTC	253	
D	b	721	TACTATTGG	GGGGAAG	AGCGGAG	CTGATTTGG	AGATAAC	CGGCTGTCA	AGTTTGATAC	780	
Q	y	254	CAG	-----	GCAG	-----	GCTCTA	TAAGTGA	-----	CCG	281
D	b	781	CAGTTGCG	CAGTTT	CAGCCAA	CGCTGTATTT	TGTGAAT	TACAAC	TAAAGCGAG	840	
Q	y	282	GGGAG	-----	CGTGGC	-----	GGTTGCAG	-----	-----	C	305
D	b	841	TGAGAAAT	TACGTAC	CCCTGG	CGCTGGA	AGTAAT	TGGTTG	MAATATGG	900	
Q	y	306	ACTGT	AGCG	-----	GACTTC	-----	-----	TTTTGGTTTCTTT	334	
D	b	901	ACTATGA	CATCGT	CACTCC	ATCCATCT	CGTCA	CATATCT	TGGTATCAT	960	
Q	y	335	-----	-----	CTC	-----	TTTTGG	-----	GCAC	347	
D	b	961	TTGAATCG	AGAGCTT	CGCGGCT	CGAGTTG	GGTGTG	ACTG	TGCTCAC	1020	
Q	y	348	-----	CTGTGG	ACTC	CCCCA	-----	GCATGA	AGGCGT	385	
D	b	1021	ACAC	CTCTGT	ATCACTA	CAACCA	ATAATTC	GCATG	CCCAAGTGTCTT	1080	
Q	y	386	G	TGCGGCGC	-----	TGCTAC	GAGCGGTGT	GTGCCT	-----	GTGCG	427
D	b	1081	GA	TGTTTCTTA	TTTTTTT	TGCTTA	TGTATTC	GCCTCG	TGTCGAT	1139	
Q	y	428	-----	-----	AGTCTG	GCCTATG	CCCGGG	-----	CCCGAGGGA	461	
D	b	1140	AGTATCT	CTACATGA	ATAAG	CACTGGT	CTCGAC	GGGAAACG	AAAAAGCGCGGA	1199	
Q	y	462	-----	CGGCGAGC	-----	TGAGAG	CGCGT	-----	GAGCTTGCTG	493	
D	b	1200	ACAA	CAGCAG	GAACG	AGATGCC	ATGTTCA	CCGCG	AGCCGAGCGCGCCCA	1259	
Q	y	494	-----	GACATGA	AC	-----	CAC	TGTACTC	-----	CTGCG	521
D	b	1260	TGCTG	ACTGT	TACTTTT	CGCGGAC	ACAAATTC	CTCTAT	GAATCCAT	1319	
Q	y	522	GGGA	ACTGGT	TACCG	-----	GAGTCC	CGAGAGGCA	-----	CTCAG	572
D	b	1320	AAAT	TGATG	TCGCG	AGATTCC	ATGATG	CAACAT	CCAGTCTGT	1379	
Q	y	573	TCCTAC	AGCGGCT	CATCG	ACTAC	TCTG	CACTCG	AGGTAG	626	
D	b	1380	ACATA	CGCTATGG	-----	CGGCTC	ATCTCG	-----	CGGCGGAA	1435	
Q	y	627	-----	CAGCGCC	CTGACCC	-----	CTGA	-----	TTGGCC	657	
D	b	1436	GC	CAACG	ATGAG	CGCTG	CAAAATCG	ATAG	CGGATAC	1495	
Q	y	658	TC	-----	CAATC	CAGAC	AGCC	-----	GAGCTCGT	694	
D	b	1496	TCT	TATCTT	CTCAAT	TATAGTCTA	CTGGTT	TATGA	ATATCTA	1555	
Q	y	695	-----	ATCTCC	-----	AACG	CAAAAG	AGCTTTT	GCCACTGA	734	
D	b	1556	AC	AGATCC	AGGAC	AGCAAGT	GCGAG	CAGATCC	ACTGATG	1614	
Q	y	735	-----	-----	GTGTCC	-----	-----	-----	TGACAC	757	
D	b	1615	AAT	CGAT	TACAA	TGGT	TACG	TGAAG	GCGGAAT	1674	
Q	y	758	AG	-----	GTGCTG	GGCG	-----	CCC	-----	787	
D	b	1675	CGG	CAAT	CGA	CTG	CGAG	TTATTA	ATCTCA	1734	


```

QY 788 -----CGGGAAC---CTCTC---CTGC-----CGAAG 809
D 1735 CAGCACTTCACTCGGGAACCTACTCTCTGCTACGGTTAGTTTCATATTTGATCGGACAG 1794
QY 810 -----CGGGAAG-----CGGGAAG 824
D 1795 CGGCTTCACTTTCTTCAATAATTTTCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1854
QY 825 -----GGGCCC---CAATCTCG---CCC-----T 842
D 1855 CTCACTTGTGATCAATCGTGAATCGGCGCTTCCGGAACCTCAATCGGTACGATGACGCT 1914
QY 843 GCCCACT-----TGAC-----TTACCA----- 860
D 1915 GCTCACTGAGACTCATCTTATGACCGGAACCAATCGACGCTTCCACCACTTGCCTATCT 1974
QY 861 -----AATCC-----CTTC-----CTGC-----AGA 876
D 1975 AAAAGCGTTGATGATATCTCTGCTGCTTCTCTGCTTATCTGCTGCTGCTGCTGCTGCT 2034
QY 877 CTAAACCTGG-----TGCTCAGGAGCGAAGGACTGTGAACCTT-----GTAGCCTGAAG-A 925
D 2035 GTAGCCTCTGTGCTCTACTCAAAAAGAGAAGAGGATCGTGGGAGAGAGAGAGAA 2094
QY 926 GCCAGAGCT-----AGCTCTG---GCCACAG-----CTGGGCGAGCTACCTT----- 966
D 2095 GACGAGCATAAACCTGCTCGGCGCACACCTGATATCTTTCAGCAGCTCGGCTTCCCGA 2154
QY 967 -----GCTCCACC-----CCACCCCAAGTTCTTAAGGCTCTTT 1000
D 2155 ATGCACATGCAACGGGCTCCACCTCGATCATCGCGCTCATCAAGCAGTCSAATCGATT 2214
QY 1001 CAGAG-----CGTGGAGGTGTGGAAGGAGTGGC-----TGCTC----- 1033
D 2215 CTGTGTCACTCACAGTCACTATGACAT---CGTACGCGTGCCTGCTTCTCTTTTCA 2273
QY 1034 -----TCCAACTATCCAAAGGG---CGGAGAGCTGCTTCTGCT---CC----- 1078
D 2274 TCTGTTCACACACTCTCTCTGCTGATTTCTGCTGATTTCTGCTGATTTCTGCTGATTTCT 2333
QY 1079 TTGAGAA-----AGGTT---CTGTTC-----CTGA---TTTATGAA---CTCTAT----- 1117
D 2334 TTAGTGAACAGAGGCTGACGCTTGGATGCTCGAGACCTTCAATTAATCTCAATCOACT 2393
QY 1118 -----AATAGATATAGGTT---TTGTAC-----CTTTTATAC----- 1149
D 2394 TCCTCATCAATTTCCATTTGCAATATCTCTTTTCTTTCGACAGAGCTTTTTCGTTT 2453
QY 1150 AGGAAGGTGACTTTCTGTAAC-----AATGC---GATG-----TATATT 1185
D 2454 TTTTATGATTTATTTTACGATTTTATAGATATGACAGATGCTCATTTGCTCAAT 2513
QY 1186 AAATTTTATAA 1198
D 2514 AAATTTTATAA 2526

```

```

RESULT 11
US-09-436-063C-3
; Sequence 3, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-3

Query Match      41.1%; Score 3097.5; DB 4; Length 2544;
Best Local Similarity 36.1%; Pred. No. 1.6e-136;
Matches 784; Conservative 0; Mismatches 404; Indels 985; Gaps 125;

QY 1 GATCTGGGGTGC-----TGCG---CA-----GGAAAAAGCAAAATTTCTGGAA 37
D 364 GATATGGACTTCATATAGACTTCTACATGGCTCAAGCGTGGCAAGCCCTCGACTAGCC 423.
QY 38 GTTAATGTTTGA-----GTGATTTT-----AATC-----CTTGTGGCGGAGAGG-- 81
D 424 TTCGGAAGTCTGATTTGGGACTTTCCAAAGAAATCGACTCATTACCGTGGAGTAGAC 483
QY 82 -----CCGCGC-----TCTCCCG----- 94
D 484 TACCTGGATAGACTGTGGAAACCGGACACGTTCTTCCAAATGAAAGAAATCATTTCTTC 543
QY 95 -----GGTATCAGGCG---TTCTCTCAT---CTTTGAATC----- 123
D 544 CACTTGGCAACACACATAACTCGTTCCTTCGTATCGAGGGTGTGGAACGGTTTATACT 603
QY 124 ---CCGCGCTCGGGT---CTTGGCGT---CA---GACGAG-----CCGGAGGAA 163
D 604 AGTCAAGATTAAACAGTCACTGCAACGCTTCCAAATGGACCTGAAAGCTGTTCCCAATGA- 662
QY 164 GCTCTTGGCAATTTAAAGC---GGGCTGTGAAAGC---CCAGGGCC---GGCGGGGGC----- 212
D 663 ---CTCTCAACTGTAAACTGMAATGAAAGCTACGGGTACGAGACGAAAGATATCGAC 720
QY 213 -----GGGCGCAGCGGGCCATTTTGAATAAAGAGG-----GTGCTTC 253
D 721 TACTATTGGGGAGAGCGGACTGATTTGGAGATAACGGCTGTCAAGTTTGATACCTTC 780
QY 254 CAG-----GCGA-----GCTCTA---TAAGTGA-----CCG---CCGCG 281
D 781 CAGTTGGCGCAGTTTCAGCAACGCTGATTTTGTGAATACAACTAAAGCCGAGACTCA 840
QY 282 GCGAG-----CGTCCGC---CGGTTGCAAGT-----C 305
D 841 TCAGAAATAACGTACGCTGCGCTGGAGTGAATATTGGTTGCAAAATATGGGCTTCTAC 900
QY 306 ACTGT-AGCG---GACTTC-----TTTGGT----- 334
D 901 ACTATGAACATGCTCATCCATCCATCCTGATCCTGATCCTCACCATATCTTGGTATCATTTGG 960
QY 335 -----CTC-----TTTGGG-----GCAC----- 347
D 961 TTGAATCGAAGACTTGGCGGCTCGAGTTGGATTTGGGTGTGACTACTGTCTCAATG 1020
QY 348 ---CTCTGACTC-ACTCCCCA-----GCATGAGGCGCT-----GAGCCCG 385
D 1021 ACAACTCTGATCACTACAAACCAATTAATTCGATGCGCAAAAGTGTCTTATGTCAGGGTCTG 1080
QY 386 G-TGGCGCGC-----TGCTACGAGCGGTGTGCTGCT-----GTG-GAAGC----- 427
D 1081 GATGTTTCTTAAATTTTGTGTTTGTGTAATGATGTTTC-GCTGTTGCTCGAGTACGCCAT 1139
QY 428 -----AGTCTGGCATCCCGCGG---GCCGAGGAGAGG---GCC-- 461
D 1140 AGTATCCTCATGAATAAAGCACTGCTTCTCGACGGGAAAAACGAAAGAAAGCCCGCA 1199
QY 462 ---CGGAGC---TGAGGAGCGCT-----GAGCTTGTGCGAC----- 493
D 1200 ACAACAGACGGAACGAGATGCCAATGTTCAACGCGAGCCCGGCAATATAA 1259
QY 494 ---GACATGAAC-----CACTGTACTC-----CCG-----CCTGC 521

```

Db 1260 TGCTGACTTGTTGCGGACACAAATTCCTCTATGATCCATGATGGAGATCCCGA 1319
 Qy 522 GGGAACTGTATCCCG---GAGTCCCGAGAGCA-----CTCAGCTTAGC-CAGGTGGA 572
 Db 1320 AAATGTGATTCGCGGACGATTCGAATGATGCAACATCCAGTCTGTGTCACAGACGGCGC 1379
 Qy 573 TCCTACAGCGCTCATCGACTACATTCCTGACCTCAGGTAG-----TCCTGGCGGAGC- 626
 Db 1380 ACATAGCTATGGC---CGGCTCATTTCGG---CGGCGGAAAGAGCTTCCAGAGATGCT 1435
 Qy 627 -CAGCCCTCGACCCG---CTGA-----TGGCG-----CCGA-----CCT 657
 Db 1436 GCCAAGATGGAGCGCGCTGCAAAATCGATAAGCTTAGCGATACGGTTTCCCATTTGCTT 1495
 Qy 658 TC-----CATCCAGACAGCC-----GAGCTCGCT-----CGGAACCTTGTC-- 694
 Db 1496 TCTCTATCTCAATATAGTCTACTGGTTGATATGAAATATCTAAGCTTAACTCGTCGG 1555
 Qy 695 ---ATCTCC-----AACGCAAAAGG-AGCTTTTCCACTGA-----C-TGGGCC-- 734
 Db 1556 ACAAGATCCAGGAGACGACAAAGTGGCAGCAGAT-CACTGATGGTATTGACGGCGCA 1614
 Qy 735 -----GTGTCC-----TGACACC--TCCAGAACGC 757
 Db 1615 AATCGAGTACAAATGGTGTACGTGCAAGAGCGGAATTGTTGACAGCGGTCAAGGCCGA 1674
 Qy 758 AG-----GTGCTGGCG-----CCC--GTTCTGCCTG-----GGACCC----- 787
 Db 1675 CGGAAACATCGAATGTCAGTTATAAATCTCAATAATCTGCCAAAGAGGACACTTGC 1734
 Qy 788 -----CGGGAAC--CTCTC--CTGC-----CGGAAG 809
 Db 1735 CAGCACTTCATCGGAGACCTACTCTCGTCTAGGGTTAGTTTCAATATGATCGGAGAG 1794
 Qy 810 -----CGGACGCG-----CGGACGCG-----CAGGGAT 824
 Db 1795 CGGCTTCTACTTCTCAATATTTTCCCTGCCAGCTCCTCGTAGTTTATCATGTAT 1854
 Qy 825 -----GGGCC--CAACTCG-----CCC-----T 842
 Db 1855 CTCATTCGTGATCAATCGTACTCGGCGCTTCGCGAACCTTAATCGGTACGATGACGGT 1914
 Qy 843 GCCCACT-----TGAC-----TTACCA----- 860
 Db 1915 GTCCTGAGACTCATCTTATGACCGGAACCAATCGACGCTTCCACCAAGTTGCTATGT 1974
 Qy 861 -----AATCC-----CTTC-----CTGG-----AGA 876
 Db 1975 AAAAGCGGTGATGATTTCTCGGTTTCTGCTATCTCTGTTTATCTGCGCTTGTATCGA 2034
 Qy 877 CTAAACCTGG-----TGCTCAGGAGGAGGACTGTGAACCTT---GTAGCTGAAG-A 925
 Db 2035 GTACGCTGTGTGCTTACTCAAAAGAAAGACGAGGATCGTCGAGAGAGAGAGAA 2094
 Qy 926 GCCAGAGCT-----AGCTCTG--GCCACCAG-----CTGGGCGAGCTCACCT 966
 Db 2095 GACGGAGCATAAACCTGCTCGCGGACACCTGATATTCTTACGACGCTCGGCTTGCCGA 2154
 Qy 967 -----GCTCCGACC-----CCACCCCAAGTTCTAAGTCTTTT 1000
 Db 2155 ATGCACATGCAACCGGCGCTCAACCTCGATCATCGCGCTCATCAACGATCGATGAT 2214
 Qy 1001 CAGAG-----CGTGGAGGTGTGGAAAGAGTGGC-----TGCTC----- 1033
 Db 2215 CTGTGTGTCAGTCACAGTCAATGACAT-CTGACGCGTGGCGGTTTCTCTTGTGTTTCA 2273
 Qy 1034 -----TCCAACTATGCCAGCGG---CGGCGAGAGCTGCTTCTGCTCT-CC----- 1078
 Db 2274 TCTGTTCAACACTCTCTTCTGCTGANTCTACTGTACAAATCAAGCGTCTGCCGTATA 2333
 Qy 1079 TTGAGAA-----AGGTT--CTGTTC-----CCTGA--TTTATGAA-CTCTAT----- 1117
 Db 2334 TTATGACACGAGGGTGACGTTGCGATGCTCCAGACCTTCAATTAATCTCAATCCAAC 2393

Qy 1118 -----AATAGACTATAGGTT--TTGTAC-----CTTTTITAC----- 1149
 Db 2394 TCCTCATCATTTTCCCATTTTGAATATCTCTTTTTCGACAGAGCCCTTTTTCGTTT 2453
 Qy 1150 AGGAAGGTGACTTCTGTAAAC-----AATGC---GATG-----TATATT 1185
 Db 2454 TTTTATTGATTTATTTTACGGATTTTAGATAATGCAAGATGCGCTCATTCCTCAAT 2513
 Qy 1186 AAACCTTTTATTA 1198
 Db 2514 AAATTTATTTAA 2526
 RESULT 12
 US-08-977-767-3
 ; Sequence 3, Application US/08977767
 ; Patent No. 5972684
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Yue, Henry
 ; APPLICANT: Greenwald, Sara
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/977,767
 ; FILING DATE: Herewith
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0423 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-855-0555
 ; TELEFAX: 650-845-4166
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1345 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 1532042
 US-08-977-767-3

Query Match 40.7%; Score 3073.5; DB 2; Length 1345;
 Best Local Similarity 46.3%; Pred. No. 1.3e-135;
 Matches 663; Conservative 0; Mismatches 423; Indels 345; Gaps 86;
 Qy 4 CTGGGGTCTGCCAGGAAAAGCAAAATCTGGAAGTTAAATGTTTTCAGTGATTTTAAA 63
 Db 8 CTGGAGGGTT---CGGAAAGGAAG---GGAAGCAAGCGGAGAGG----- 52
 Qy 64 TCCTTGCTGGCGAGAGCGCCCTCTCCCGGTATCAGCGTTCCTCAT-TCCTTGAAT 122

Db 53 ---AGAGAGAGAGAGGGGTGGGGTCCAG--ATCCCGCTGCTGCTGTAACCTGGGAA 107
Qy 123 CCGCGGCTCC-----GGGTCTTCGCG-GTACAGACAGCCGGAGGAAGCTGTT 170
Db 108 CTGGGACTCCCTCAGCAGAGAGCG--AAGCTGGAAGTCTTG--GGAGGATTAAGGT 162
Qy 171 TGCAATTTAAGCG-----GCTGTGAACGCCAGCGCGCGGGG-----GGGCG 217
Db 163 TCCACAAGAGGGGGAAGCTGAGAG-----AGGGCCGGCCGGGACATGTGGGAAGGGGT 217
Qy 218 CGAGGGGGGCAATTTGAATAAGAGCG-GTGCT--TCCAGCAGGCTC-----TATA 268
Db 218 CACCCCGAGGCTCTTTGGGATGGGGCTGCACCTGCTGAGC--GCCCTCGGGTGT 275
Qy 269 AGTGAC--GCCGCG--GGGAGCGTGGCGGCTTGA--GGTCACTGTAGCGACT--- 318
Db 276 GGTACTCTGGCGCGCACTGGGAGCG--GCACGTCACTCGGACCC--GCACCTGACCA 331
Qy 319 --TCTTTTGT--TTTC-----TTTCTCTTTGGGCACT-----CTGACTCACTCCC--- 363
Db 332 GGACTGTGAGCTACAGGATTAATCTCAGGGAACTTGTGCCAGGAC---CTCCCTTC 388
Qy 364 -----CAGCATGAAGCGCGTGAAGC--CGGTGCGGCG--TGCTACGAG 404
Db 389 TGGGGCTGGTGAACGAGCGCTGAGTCTCTGTGCGCTGGGAAGCGTCAAGCGCCGCTG 448
Qy 405 CGGTGTC--TG-----CCTGCGAAGCGAGTCTG--CCATC-----GCCGCGGCG 450
Db 449 ATGTGGACGTGAAGAGGGTCTCTATGACCCCTTCTGTGCCCTCTGAGACTCAGCACG 508
Qy 451 AGGGAAGGCGCGCAGCTGAGGAGCGCTGAGCTT--GCTGGAACGATGAACCACTGCT 509
Db 509 GGGAG--AAGCCTCGTGAACT--CTGTACAACACCGGTGCGCATG-----TCTCCT 559
Qy 510 ACTCCCGCTGGCGGAACGTGGTACCGGAGTCCGAGAGGCACTGACTAGCCAGTGG 569
Db 560 --TCCTGCTTGG-----CCCGCG--CCTGTG-----TTAATGTGTGCGGGG 600
Qy 570 A-AATCTCAGCGCGTCACTCACTACTCTCGAC--CTGC-----AGGTAGTCTCTGGCC 622
Db 601 ACCTCTTTATAG--CACGACTCAGTGAACCTGCGGCTGCTATTGCGGAGCAGGAG 658
Qy 623 GAGCAGCCCTGGAACCCCTGATGCGCC-----CAC-----CTTCCAT 663
Db 659 GAGTGTGCTCT--GAACACAGATCAACCATCAGGGTCTCTGCTGAGGTGCACTCAT 716
Qy 664 CCA-GACAGCGAGCTCGCTCGGAACCTTGTCTATCTCCACGACAAAGAGGCTTTGCG 722
Db 717 CCACTTCAACAGAACTCTATGGA-----ACCTCAGTGGCGCCACGAGGCG-----CC 768
Qy 723 ACTGACTCGCGGTGCTCTGACACC-----TCCAGAACGAGGT--GCTGGCGCGGTTTC--- 775
Db 769 AATGGCTGGCACTTCTCAGCCTTTTGTCAATGTGGCTGGTAGCTCAAAACCGTTCCTC 828
Qy 776 ---TGCCCT-----GGGACCCCGGAACC--TCTCCT-----GCCGGAAGCGG--A 814
Db 829 AGAGCCTCTTAACTGTGACACCATCACCAGATCTCCTATGAAGATGATGCTACTTT 888
Qy 815 CGGAGGG--ATGGGCGC-----CAACTTCGCGCTGC--CACTTCACTTCACCAATCCCTT 868
Db 889 CTTCAAGACCTGAGCCTGGAGCTCTGTGCGCCGAGTCTCTTGGCTTCATCA-----CCTA 944
Qy 869 CTTGG-----AG-ACTAACTGTGCTCAGGAGCGAAGGACTGTGACTGTGAGCC 919
Db 945 TCAGGGTCTCTAGCAACCCACCC--TGCTCG-----AGACTGT--TACCTGATGTC 994
Qy 920 TGA-----AGAGCCAGAGTACTGTGGCCACAGCTGGGCGACGTCACTGCTCCAC 974
Db 995 TCAITGACAGGGCCCTCAATATCACTCCCTCCAGCTG-----CAT--TCCCTGCACTTC 1048
Qy 975 C---CCA-----CCCCGAAGTCTTAAGTCTTTTCAAGCGGTGAGGTGT--GGAGG--AGT 1025
Db 1049 TGAGCCGAATCTCTCGGGTCCCA--GTC--TTCCAGAGCTCAGCCGCTAAAGCGCGCGCC 1105

RESULT 13

US-09-627-650B-5
; Sequence 5, Application US/09627650B
; Patent No. 6406872

GENERAL INFORMATION:

; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1917
; TYPE: PR
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-5

Query Match 39.4%; Score 2971.5; DB 4; Length 1917;

Best Local Similarity 39.7%; Pred. No. 8.6e-131;

Matches 729; Conservative 0; Mismatches 435; Indels 671; Gaps 108;

Qy 3 TCTGGGCT--GCTGCCAGGAAAAAGCAATTCCT---GGAAGTT-----AATGG 45
Db 100 TCTGTGCTCCCGCTAACCG--ATTGGCTCAATTCCTTCGCTCTTGTCACTAACGAATCG 157
Qy 46 TTTT-----GAGTGATTT---TAAATCCTTGC-----TGG-----CGGA--- 77
Db 158 CTTTTCACAGAGTGTATAGGAGAAAAACCTCCCAACATTGGCTCACACCCGATTA 217
Qy 78 -----GAGG--CCGCGCTCTCC----- 92
Db 218 TGAATCTTCTGCTGCTCTGCTGCTCTCTGCTGTAGTGTAGACGAAGAAGAAGAAG 277
Qy 93 --CGGTATC-----AGCGCTTCC--TCATTTTGAATCCGCGCTCGCGGTCTTC 141
Db 278 CTCATTTCTCGAAGAAATGGCTCGTCCATTACACTTATCGTACTCTCTCCGC--ACATC 335
Qy 142 GCGCTC-----AGAC-CAGCGGAGGAAGC----- 165
Db 336 TGTGTCTACATGTGGTGTGACACAGGATGAGGACTCACATATCAACTCACTCCTCT 395
Qy 166 ---CTGTTTGAAT-----TAAGCGGCT-----GTGAAGCCCGAGGCGCGG 208
Db 396 CATCAGTTCTCGATAGACTCAGGAATCGCACTACTTATGATAAAGATTACGGCC--CAG 453
Qy 209 GGGCGGGCCGAGGCGGCGCCAT--TTTGAATAAAGAG--GGTGGCTTCCA---GGCAGG 261
Db 454 GTATGGTGAAGCCAGTCGACGTGGATTTACATACACGTTTCTTCAATCTCTGCAGT 513

396	DB	CATCAGTCTTCGATAGACTCAGCATCGCACTACTATGATATAAAGATTACGGCC--CAG	453
209	QY	GGCGGGGGCGAGCGCGGCCAT-TTTGAATAAAGAG--GCGTGCCTTCCA---GGCAGG	261
454	DB	GTATGCTGAAAGCCAGCTCGAAGTTGGAAATACGATACACAGTTTCTTCAATCTCGAGT	513
262	QY	CTCTATAAGT-----GAC--CGCC-----GCGCGAGCGTGCOCGCTTGCAGTTC	305
514	DB	TTC-AGAAGTTGATATGGACTTCACATATAGACTTCTACATGGCT-CAACAGTGGCAAGAC	571
306	QY	ACTGTATCGGACATCTTTTGGTTTCTTCTCTTTTGGGC--ACCTCTG-----GACTCA	358
572	DB	CC1-----CGACTAGCCTTCGGAAGTCTT-GATTTGGGACTTTCCAAAGAAATCGACTCA	625
359	QY	CTCCCCAGC-ATGAAGC-----GCTGAG--CCCGGTGCG-----CGG--	393
626	DB	CTTACCGTCGGAGTAGACTACCTGGATAGACTGTGGAACCCGACACGTTCTTCCCAAT	685
394	QY	-----CTGCTACGAGGG-----GTGTGCTCGCTGTC-----	420
686	DB	GAAAGAAATCAITCTTCACATTGGCAACACACATAA CTCGT-TCCTTGGTATCGAGG	744
421	QY	-----GGAACG-----CAGTC-----TGGCATCGC-----CCGGGCGGAGGAAGG	458
745	DB	TGATGGAACGGTTTATATAGTCAAAGATTAA CAGTCACTGCAACGTGTCC-----AATG	799
459	QY	GCCCGCAGCTG-----AGGAGCGCTGAGCT--TGCTGACGACATGAACCACTGC	508
800	DB	GACTGAAGCTGTTCCCAATGGACTCTCAACA CTGTAACTGGAATTGAAAGTACGGC	859
509	QY	TAC-----TCC-CGCC-----TG-----CGGAA-----CTGGTACCCGGAGTC-	541
860	DB	TACAGTATCTCGACATTATGTACGTGTGCGACGAGAAGAA GTCCGTGTCCACCGAGTCT	919
542	QY	-----CCG-----AGAG-----CACTC-----AGCTT	559
920	DB	TATGAGTTCGCGAGTTTGTACTTCACTATCAAGTCTGCTCAATCATACGCAAAAGCTT	979
560	QY	AG-CCAGGTGGAAA-----TCCTA---CAGCGGCTATCGACTAC	595
980	DB	AGITCAGGAATATCCCGCCTTTGCTGTTCTTCTCTTA TTCAAGCGTAACATCGGCTTC	1039
596	QY	ATTCTCGAC-----CTGCAAGTAGTCTCTGCGCG-AGCAGGCCCTGGACCCCTCGA	645
1040	DB	TACATCATCAAAATATATCTACCATCTGCTCTGA TTGCTGCTATCTCATGGGTATCTTTT	1099
646	QY	TG-----GCCCC-----CACCTTCG-----CATCCAG-----ACAGCCGAGCTCGCT	682
1100	DB	TGTTTGAGCCGGATGCGACCGCGCAAGAGTTGCTCTCGAGTTCACCACTGTGCTACA	1159
683	QY	CCG-----GAACTGTCTATCTC---CAACGCAAAAG-----	711
1160	DB	ATGACTACTTTGATGACCATGACTAATAGTTTCAATGCGCAAAAGTGT CATATGTGAAAAGT	1219
712	QY	---GAGCTTTTGCA-----CTGACTC--GSCCGTGTCTTG-ACACCTCCAGA--ACGCA	758
1220	DB	ATCGATATATTTCTAGGTGTCGTCTCATGATGTATCTGTTCACCTCTTAGAATAGCC	1279
759	QY	G--GT-----GCTGGCGCCCGTTC-TGCGTGG-----	782
1280	DB	GCCGTTCGGATACATACGCAAAACGATGAAGCTTTCGCGAGCGCAAAAAGAAATCTCGAATG	1339
783	QY	--GACCCCGGAAC-----CTCTCTCGCGGAGCGGACCGCAGGAGTGGCC	829
1340	DB	CTGACCCCTTTACCAATCTTGAAGTCTCTTCTCTCC--AAACGTTACTCTA--TCCGTTC	1395
830	QY	C-----CAACTTCGCC-----CTGCCCACTTGACT--TCACCAATC-----CCTT	868
1396	DB	CTCGTATTTCAACACACCAGTACCGCCGCTTTTACTGTCTCCACCGATCAACGTCCAA	1455
869	QY	CCTGAGACTAAACCTGTGTCTCAGGACGGAAGACTGTGACT---TGTAGCTGAAG	924
1456	DB	CCTGTACATT---CCGAGTCCG---ACGCACACAGATTTTCTCAAATG-AGATGCG	1508

RESULT 15

US-09-627-650B-1

US-93-027-000B-1
: Sequence 1, Application US/09627650B

Patent No. 6406872

FACEBOOK NO: 0400072
: GENERAL INFORMATION:

APPLICANT: Bamber, Bruce

REFUGIAN: Jorgensen Erik
APPLICANT: Jorgensen Erik

TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and

TITLE OF INVENTION: Nematode Neuroanatomical

; TITLE OF INVENTION: MECHANISM
 : FILE REFERENCE: 21101-0009UE

FILE REFERENCE: Z101.000003
CURRENT APPLICATION NUMBER: US/09/627.650B

; CURRENT BILLING DATE: 20

; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063

;; PRIOR APPLICATION NUMBER: 09/436,063
: PRIOR FILING DATE: 1999-11-08

; PRIOR FILING DATE: 1999-II-08
 : PRIOR APPLICATION NUMBER: 60/107.727

; PRIOR APPLICATION NUMBER: 80
: PRIOR FILING DATE: 1998-11-06; PRIOR FILING DATE: 1998-
; NUMBER OF SEC ID NOS: 50

```

; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1

```

```

; SOFTWARE: PAC
; GEO ID NO 1

```

```
; SEQ ID NO 1
:
: LENGTH: 1652
```

```

;
; LENGTH: 1652
; TYPE: PPT

```

TYPE: PRI
ORGANISM: Caenorhabditis elegans

;; ORGANISM: Ca
HG-08-E27-CE0B-1

Overall Match	39 2*	Score	2954.5	DB 4	Length	1652;
---------------	-------	-------	--------	------	--------	-------

Query Match

Best Local Similarity	41.6%	Pred. NO. 4.7e-130;	Indels	497:	Gaps	98:
Matches	580:	Conservative	Mismatches	458:		

1 CATTCCCTCCCTCCCTC - CCACCAAAACGCAATTCTC--GAAGTTAATGTTTGAGTGAT 56

1 GATCTGGGGTGCCTG--CCAGGAAAGCAAAATCTG--GAAGTAAATGGTTTGGTGAATGG

b
40 GATCT--TCTGCTGCTCCTG---CTGGCTCCCTCTGCTGTAGTIGA-GACGAAGAAGAAAG 92

----- 92

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

b
93 AAGAAGCTCCATTCTCGAGAAATGGCTCGTCCATTACACACTTATCGTACTCTCTCCGCA 152

93 - - - - - CCGG - - - - - TATCAGCGCT - - - - - TC 108

[illegible]

Qy 139 -----TTGGCGTCAGACAGCGGAG-----GAAGCTGT-----TGC 173
 Db 272 CAGGATATGTAAGACCGAGTCGAGCTTGGAAATACGATACACGTTCTTCAATCTCTGC 331
 Qy 174 AATTTAAGCGG--GCTGTGAAC--GCCA--GGGC-----CGCG--GGGCGGG 216
 Db 332 AGTTTCAGAAGTGTATGGACTTCAATAGACTTACATGCTTCAAAAGTGGCAAGA 391
 Qy 217 -CGAGCGGGCC-----ATTTTGAATAAGAGCGTGCCTTCCAGCAGGCTCTATA 268
 Db 392 CCTCTGACTAGCTTCGGAAGTCTTGATT--TGGACTTTCAAAAGAAATCGACTCACTT 449
 Qy 269 AGTGACCGCGCGCGAGCGTGGCGCGT-----TGCAGTCACTAGCGGACTT--CTT 322
 Db 450 ----ACCGTCGGAGTAGAC--TACCTGGATAGACTGTGGAAACCCGACAGCTTCTCCCAA 504
 Qy 323 TTGGT-----TTTTCTTCTTTTGGGCACTCTGGACTCACTCC-----CGAGC 367
 Db 505 ATGAAGAAGAAATCAATTTCTTCCACTTGGCAACACACATAACTCGTTCTTCTGATCGAGG 564
 Qy 368 ATGAAGCGCGTG-----AGCC--CGGTGCGCG--CTGCTAGGAGC-----GG--- 407
 Db 565 GTGATGACCGGTTTATAGTAGTCAAGAGATTACAGTCACTGCAAGCTGCAATGGACC 624
 Qy 408 TG-TGCTG---CC--TG--TCGGAACGCACT---CTGGC-----CGAGGCGC 435
 Db 625 TGAAGCTGTTCGAATGAGACTCTCAACACTGTAAACTGGAAATGGAAGCTATGCGTATT 684
 Qy 436 CATCGCCCGGGCGGAG-----GGAAGGCGC-----CGGAGCTG 470
 Db 685 CGAGCGCGGAAATCGAGTACAAATGTGTACGTGCGAAGGAGCGGAATTTTCGACAGCGG 744
 Qy 471 AGGA--GGCGCTGAG-----CTTGTGGAGACATGAAC---CACTGCTA--CTCCCGCCTG 520
 Db 745 TORAGCGCGAGCGGAACATGMACT--GTGAGTTTAAATTCATTAATCTGCCAATAA 803
 Qy 521 CGGGAACCTGGTACCGGAGTCCGAGAGGCACTGAGCTTAGC--CAGGTGGAATCTCTAC 578
 Db 804 CGGACACT--TGCCAGCACTTCATCGGGGACCTACTCTGCTACGGGTAGTTTCATAT 861
 Qy 579 ---AGCGCTCATCGATACATTTCTCGACCTGAGGTAGTCTCGCGGAGCGGAGCC--- 631
 Db 862 TTGATCGGACAGCGGCTTC--TACT--TTCTTCAATATTTT---CCCTGCCAGCCTCGT 916
 Qy 632 -----CTGGACCCCGCTGATGGCCCC-----CACTTC--CCATCC-- 665
 Db 917 CGTAGTTTATCATGGATCTCATTTCTGATCAATCGTGACTCGCGGCGCTTGGGAAACCT 976
 Qy 666 -----AGACAGCGAGCTCGCT-----CGGAACTTGTCTATCTCCA 701
 Db 977 AATCGGTACGATGACGGTGTCTACTGAGACTCATCTTATGACCGGAACCAATCGAGTCT 1036
 Qy 702 ACGACAAAGGAGCTTTT-----GCCACTGA-----CTCGCGCGTGTCTCTGACACCT 748
 Db 1037 TC--CACCAGTTGCTATGTAAGCGCTTGATGATATTCCTCGT---TTCTGCTATCT 1090
 Qy 749 CCAG--AAGCGAGTGTG-----GGCCCGCTTCTGCTGGGACCC----- 787
 Db 1091 TCTGTTTACTGGCGTTGATCGAGTAGCGCTGTGTTGCCT---ACTCAAAAAGAGAA 1147
 Qy 788 CGGGAACCTCTCCTCCCGAGCGGAGCGGAGCGGAGGATGG--GCCCAACTTCGCCCTGCC-- 845
 Db 1148 CGAGGATC-----GTGCG--AGAGAGAGAGAGAGCGGAGCATAAACCT--GCTCCGCG 1199
 Qy 846 -CACTTGA-----CTTCAACAATC--CCTTCTGAGACTAACT--GGTGTCTAG--- 893
 Db 1200 ACACCTGATATTTCTACAGAGCTCGCGCTTCCGAATGCAATGCAACGCGGCTCCAACC 1259
 Qy 894 --GAGCGAGGAGCTGTGACTGTAGCTGAA--GAGCCAGAGCTAG--CTCTGGC----- 943
 Db 1260 TCGATCATCG--CGGTATCAGGAGTTC--GAATCGATCTGTGTGCTCAGTCACATTCATT 1316

Search completed: May 7, 2004, 15:11:36
 Job time : 58 secs

Qy 944 -----CACCAGCTGGCGAG-----TCACCGTGTCCACCCACCCACCC--- 982
 Db 1317 GACATGTCAGCGTGCAGGTTTCCTCTTGTGTTTCTTGTGTTTCAACACTCTCTTCTGG 1376
 Qy 983 CAAGTTCTAAGGT--CTTTTCAGAGCGT--GGAGGTGTGAAG-----GAGTGGCTGC 1031
 Db 1377 CTGATTTCTACTGTACAAATCCAAAGCGTCTGCCGTATATTAGTGAACACGAGGCGTACCGT 1436
 Qy 1032 T-----CTCCAAC-----TATGC--CAAGGCGGCGAGAGCTGGTCTTCTGGTCTCC 1078
 Db 1437 TGGATGCTCCAGACCTTCAATCTCAATCCAACTTC-----CTCATCAT---TTTCC 1488
 Qy 1079 TTGGAGAAAGGTTCTG-----TTGCCCTGATTTATGAACCTCTATAATAGAGTATATAGTT 1134
 Db 1489 ATTTGGAATATCTCTTTTCTTGACAGA-----AGC--CTTTTTTCGTTTTTTTT--TA 1539
 Qy 1135 TTGTACCTTTTACAG-----GAAGGTGACTTTTCTGTAACAATGCGAGTATATATA 1187
 Db 1540 TTGATTTATTTTACGGATTTTAGATAATGCAAGTGCCTCATTC-----TCAATATA 1595
 Qy 1188 ACTTTTTATAAAGT 1202
 Db 1596 ATTTATTTTAAATGT 1610

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:10:05 ; Search time 54 Seconds
 (without alignments)
 6183.570 Million cell updates/sec

Title: X69111
 Perfect score: 7544
 Sequence: 1 GATCTGGGCTGCTGCAGGA.....TTAAACTTTTATAAAGTT 1203

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*	Sequence 543, App
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*	Sequence 543, App
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*	Sequence 543, App
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*	Sequence 543, App
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*	Sequence 543, App
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*	Sequence 543, App
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*	Sequence 543, App
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*	Sequence 543, App
9:	/cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*	Sequence 533, App
10:	/cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*	Sequence 533, App
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*	Sequence 517, App
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*	Sequence 517, App
13:	/cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*	Sequence 405, App
14:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*	Sequence 405, App
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*	Sequence 405, App
16:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*	Sequence 405, App
17:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*	Sequence 405, App
18:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*	Sequence 541, App

ALIGNMENTS

RESULT 1
 US-10-142-426-381
 ; Sequence 381, Application US/10142426
 ; Publication No. US20040048333A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: Deforse, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C224
 ; CURRENT APPLICATION NUMBER: US/10/142,426
 ; CURRENT FILING DATE: 2002-05-09
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 381
 ; LENGTH: 2236
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-142-426-381

Query Match 47, 68; Score 3593; DB 12; Length 2236;
 Best Local Similarity 39.88; Pred. No. 6.2e-171;
 Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	3593	47.6	2236	12	US-10-142-426-381
2	3593	47.6	2236	14	US-10-123-155-381
3	3593	47.6	2236	14	US-10-146-731-381
4	3593	47.6	2236	14	US-10-140-472-381
5	3593	47.6	2236	14	US-10-141-761-381
6	3593	47.6	2236	14	US-10-142-885-381
7	3593	47.6	2236	14	US-10-158-790-381
8	3593	47.6	2236	15	US-10-137-871-381
9	3593	47.6	2236	15	US-10-140-923-381
10	3593	47.6	2236	15	US-10-141-756-381
11	3593	47.6	2236	15	US-10-141-759-381
12	3593	47.6	2236	15	US-10-140-805-381
13	3593	47.6	2236	15	US-10-140-864-381
14	3576.5	47.4	3721	12	US-10-142-426-543
15	3576.5	47.4	3721	14	US-10-123-155-543

QY	1	GATCTGGGGTGTCTGCAGGAAAAAGCAAAATTC--TGGAAGTTAATGTGTTTTGA-----GTG 54
Db	191	GGCCGCGGGGAAACCGG-----GCGGAATTCCTCGCGGTCAAAACCACTGATCCCAT 244
QY	55	ATTTTTAAATCCCTTCTCTGCG-----GAGAGGCCCGCCTCTCCCGGTATCAGC 103
Db	245	AAACATTCATCTC-CCGGCGCCGCGCTGCGAGCGCCCGCACTCGCGGCCGCGCC 303
QY	104	GCTTCTCTCATTTTGAATCC-GCG-GCTCGCGGTCTTGGCGTC--AGACCAGCGG 158
Db	304	GC--CTTCGCCCTGTGCGCCCTGCGCCCTTGG--CACCGGGGCCGAGCCAGCCAG 359
QY	159	AGGAAGCTGTGTTGCAATTTAAGCGGGCTGTG-AAAG-CCCAAGGCC-GGCGGGGGGGG 215
Db	360	AGCGGGCGG--AGCG-----GAGGCGCCGAGCCTGTTCGCGGCCGCGCGCGGCGG 413
QY	216	GCCGAGCGCG-----GCCATTTTGAATAAAG--GGC-----GTGCTTTCCA 255
Db	414	GCCGTAGCGCGCGCC-----TGATGCGGAACCGCGCGGGGGAAGCGGCGCCCGCC 469
QY	256	GGCAGGCTCTAAGTGAACCGCGG-----GCGAGCGGTGG 292
Db	470	CGAAACGACTTTTCACTCCCGACGCGCCGCCAACCCCTACGATGAAGAGGGCGTCCG 529
QY	293	C-----GC-GTTGCAGGTCA-----CTGTAGC---GGACT-TCTTTTGTGTT--T 329
Db	530	CTGGAGGGAGCGGCTGTCTGG-CA TGGGTGCTGTGCTGCGGCCTGCGAGCTGCGAGC 588
QY	330	TCTTTCTCT--TTGG--GCACCTCTGCACTCA-----CT-CCCCAG 366
Db	589	CCATGCCCAGGTGCTGTGATGTGTA CAATGAGCCCAAGGTGACACAAAGCTGCCCCAG 648
QY	367	CA-----TGAAGCGCTGAGCCGG-GTCCGCGGTGTACAGAGGGGGTGTG-CTGCTG- 418
Db	649	CAGGGCTGCAAG-GCTGTGCCGTGGGCATCCTGCTGCCAGCCAGCGCATCTTCCTGC 706
QY	419	TGGAA-----CGCA--GTCTG-GCCATGCCCGGG---GCCGAG-----452
Db	707	ACGGCAACCGCATCTCGCATGTGCAGTGCAGCTTCCGTGCTGCCCAACTCACC 766
QY	453	---GGAAG-----GGCCG-----GACGTG-----470
Db	767	TCCTGTGGTGCACCTCGAATGTGTGGCCGGAATTGATCGGTGCTTCACTGCGCTGG 826
QY	471	---AGGAGCGGTG-----AGC---TTGC-----TGG-----491
Db	827	CCCTCTGGAGCAGCTGACCTCAGCATATGCAAGCTCCGGTGTGTGACCTGCCA 886
QY	492	---ACGACATGAACAC-----TGCTACTCC-----C-GCTTGGGGGAC 527
Db	887	CATTCCAGGCTGGGGCCGCTACACACGTGCACCTGGACCGCTCGCGCTTCGAGAGC 946
QY	528	TGTTACCCCGAG---TCC-----C 543
Db	947	TGG-GCCCGGGCTGTTCCGGGCGTGGCTGCCCTGCAGTACCTCTACCTGCAGAGAAC 1005
QY	544	GAG---AGGCACT-----CAGCTT---AGCCAG-----566
Db	1006	GCCTCGAGGCACTCCCTGTATGACCTTCGGGACCTGGGAACTCACACCTCTTC 1065
QY	567	---TGGA---TCCTAC-----AGCGGTATC-----GACT--ACATTTCTC 601
Db	1066	CTGCACGGCAACCGCATCTCCAGCGTGGCGAGCGCGCTTCGTGGGTGACAGCTC 1125
QY	602	GAC-----CTGCA-----GGTAGTCC-----617
Db	1126	GACGCTCTCTACTGACACAGACGGGTGGCCCATGTGACCCGGATGCCCTTCGGTAC 1185
QY	618	--TGGCG-----AGCCAG-----630
Db	1186	CTTGGCGGCTCATGACACTCTATCTGTTTGGCAA CAATCTATCAGCGTGGCCACTGAG 1245

QY	631	CCCTGACCCCTGTATGGCC--CC--ACCT-----TC-----CCATCC-----A	666
Db	1246	GCCTTGGCCCCCTGGTGCCTGCACTGAGGCTCAACGACAA	1305
QY	667	GACAGCCGAGCTCG-----CTCCGGAATTGT-----CATCTCCAAACGACAA	709
Db	1306	GACTGCCGGCAGCCCACTCTGGCGTGGCTGCAGAAATTCCGCGGCTCTCTCTCCGAG	1365
QY	710	AGGAGCTTTTGGC-----ACTGACTCGGCCGTGTCCT--GACACCT-----CCA-	751
Db	1366	GTGCCCTGACGCTTCCGGAACGCTTGGCT--GGCCGTGACCTCAACGCCCTAGCTGCCAA	1424
QY	752	GAA-C-GCAGG--TGC-----TGAG-----GCCCGT-----TCGTG--C--CTGGGA	784
Db	1425	TGACCTGCAGGGTGTGCCTGTGGCCACCGGCCCTTACCATCCATCTGGAACCGCAGGC	1484
QY	785	CCCGG-----GGAACTCTCT-----CCTCCGGAAGCG--GAC--GG--GG--817	
Db	1485	CACCGATGAGGAGCCGCTGGGGCTTCCCAAGTGTCTGCCAGCCAGATGCCGCTGACAAGC	1544
QY	818	--CAGGATGG--GCC-----CCAACTTTCG-----838	
Db	1545	CTCAGTACTGGAGCTGGAAAGCACGCTTCGGCAGGCAATGCGTGAAGGACGCGTGCC	1604
QY	839	--CC-----CTGGCCACTTGACTTC-----ACAAATCCCTTC--T	871
Db	1605	GCCCGGTGACAGCCCGCGGGCAACGCTCTGGCCCAACGGCACATCAATGACTCACCCCT	1664
QY	872	GGAGCTAAACCTGG-----TGCTCA-----GGAGCGAAGG-ACTGT	907
Db	1665	TGGGACTCTCGCTGGCTCTGCTGAGCCCCGCTCATCTGAGTGGCGCCGAGGGCTCCGA	1724
QY	908	G-----AACTTG-----TAGCCTGAAGAGCCAG-----AGCTAGCTC	939
Db	1725	GCCACGAGGTTCGCCACCTCGGCGCTCGCCGGA--GGCCAGGCTTTCACGCAAGAAC	1782
QY	940	TG-----GCCA--CCAGCTGGG--GAGG-----GAGG-----959	
Db	1783	CGACCCGACGACCATGTCCTCTGGGCAGGACGAGGCGGGGTGGGGACTGTGTAC	1842
QY	960	-----TCA-----CCCTGTCT-----CAACCCACCC--982	
Db	1843	TCGAAGGCTCAGGTGCTTACCCAGCCTCACCTGAGCCTCACCCCTGGGCCCTGGCG	1902
QY	983	CAAGTCTTAAGTCTTTTC-----AGAGCGTG-----1009	
Db	1903	CTGGTGTCTGGACAGTCTTGGCCCTGTGACCCCCAGCGACACAGAGCGTCTCA	1962
QY	1010	-----GAGGTGTG-----GAAAGGATGGTGTCTTCCAAATATGCCAAGGC-----GG	1053
Db	1963	GCAGCCAGGTGTGTGTACATACCGGGT--CT-CTCTCCACGC--CGCCAAAGCAGCCGGG	2017
QY	1054	CGGCGA-----GCTGTCTCT--TCTGTCTCTCTGGAGAAAGGTCTGTGTGCCCTG	1102
Db	2018	CGGCCGACCGTGGGGCAGCCAGGTCCTCTCTGTATGACG--CCTGCCGCCCGC	2075
QY	1103	ATTATGAACTCTA--TAATAGAGTATATAGTTTGTG-----TACCTTTTTCACGAGAG	1155
Db	2076	CACCCCATCTCACCCCATCATGTTTACAGGGTTGGGGGACGCGTTGTTCAGAACG	2135
QY	1156	GTGACTTTCTGTAAACATGCGATGTATA-----TTAAACTTTTATAAAGTT	1203
Db	2136	CGCGCTCCCAACCCAGATCGGGTATATAGAGATATGATCATTTATTTTACTT	2186

RESULT 2
US-10-123-155-381
; Sequence 381, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforse, Laura


```

; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-123-155-381

```

```

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171; Indels 863; Gaps 125;
Matches 808; Conservative 0; Mismatches 360;

```

```

QY 1 GATCTGGGGTGTCTGCGAGGAAAGCAATTC--TGAGATTAAATGGTTTGA-----GTG 54
DB 191 GCGCGCGGGGAAACCG-----GCGATTCTCTCGCGGTCAAAACCACTGATCCCAT 244
QY 55 ATTTTAAATCTTGTCTGGG-----GAGAGCGCGCTCTCCCGGTATCAGC 103
DB 245 AAACATTATCTTC-CGCGCGCGCGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCG 303
QY 104 GCTTCTCTATCTTTGAATCC-GCG-GCTCCGCGGTCTTCGGGTC---AGACCAAGCGG 158
DB 304 GC--CTCGCGCTGTGGCGCTGCGCGCTGCG--CACCGCGCGCGCGAGCGCGCGCG 359
QY 159 AGAAGCCTTTTGAATTTAAGCGGGGTGTG-AAAG-CCAGGCGC-GGCGGGGCGGG 215
DB 360 AGCGGCGCG--AGCG---GAGCGCGCGAGCGCTCGTCCGCGCGCGCGCGCGCGCG 413
QY 216 GCGAGGCGG---GCCATTTTGAATAAGA---GGC-----GTGCTTTCCA 255
DB 414 GCGTACGCGGCGCG---TGATGCGACCGCGCGCGCGGAGACGCGCGCGCGCGCG 469
QY 256 GCGAGGCTTATAGTACCGCGCG-----GCGAGCGTGGC 292
DB 470 CGAAACGACTTTCAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
QY 293 C-----GC-GTTGCGAGTCA-----CTGTAGC---GGACT-TCTTTTGGTT--T 329
DB 530 CTGAGGAGCGCGCTGCTG--CATGGGTGCTGTGTGCTGAGCGCTGCGAGGTGGCAGCC 588
QY 330 TCTTTCTCT-----TTGGG---GCACCTCTGGAATCA-----CT---CCCGAG 366
DB 589 CCATGCCCGAGTGCCTGGTATGCTACAAATGAGCCCAAGGTGACGACAAAGCTGCCCGCG 648
QY 367 CA-----TGAGGGCTGAGCGCG-GTGCGGGGTGCTACGAGCGGTGTG-CTGCTG 418
DB 649 CAGGCGCTGCAG--GCTGTGCGCGTGGGATCCCTGCTGTGCGAGCGCGATCTTCTGTC 706
QY 419 TCGGAA-----CGCA---GTCTG--GCCATCGCGCGG---GCCGAG-----452
DB 707 ACGGCAACCGCATCTGCATGTGCAGCTGCGAGCTTCCGTGCTGCGCGAACTCACCA 766
QY 453 -----GGAAG-----GGCCG-----GCAGCTG-----470

```

```

DB 767 TCCTGTGGCTGCACTCGAATGTGTGGCCCGAATTGATGCGGTGCTTCACTGGCGCTGG 826
QY 471 -----AGAGCGCTG-----AGC-----TTGC-----TGG-----491
DB 827 CCCTCTGTGAGAGCTGGACCTCAGCGAATAATGACAGCTCCGGTGTGTGGACCTTGCCA 886
QY 492 -----ACGACATGAAACAC-----TGCTACTCC-----C-GCTTGGGGAAC 527
DB 887 CATTCCACGGCTGGCGCGCTACACACGCTGACCTGACCGCTGCGGCTGCGGAGC 946
QY 528 TGTATCCCGGAG-----TCC-----C-----C-----C 543
DB 947 TGG-GCCCGGGGTGTTCGCGCGCGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1005
QY 544 GAG-----AGCACT-----CAGCTT-----AGCCAGG-----566
DB 1006 GCGCTGCGAGCACTGCTGATGACACTTCCGGGACCTGGGCAACCTCACACACCTCTTC 1065
QY 567 -----TGGAAA-----TCTTAC-----ACGGGTGATC-----GACT--ACATTCTC 601
DB 1066 CTGCACGGCAACCGCATCTCCAGCGTGCAGCGCGGCTTCCGTGGGTGCGACGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC-----617
DB 1126 GACGCTCTCTACTGACACAGACCGGCTGCGCCATGTGACCGCGATGCTTCCGTGAC 1185
QY 618 -----TGGCGG-----AGCCAGC-----630
DB 1186 CTTGGCGCTCATGACACTCTATCTGTTTGGCAACAATCTATCAGCGTGCACCTGAG 1245
QY 631 -CCCTGACCCCTGTATGGGCC--CC-ACT--TC--CCATCC-----A 666
DB 1246 GCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1305
QY 667 GACAGCGAGCTCG-----CTCGGAACTTGT-----CATCTCAACGACAAA 709
DB 1306 GACTGCGCGGACCGCCACTCTGGCGCTGCTGCAAGTTCCGGGCTCTCTCTCGAG 1365
QY 710 AGGAGCTTTTGC-----ACTGACTCGGCGGTGCT--GACACCT-----CCA- 751
DB 1366 GTGCGCTGAGCGCTCCCGCAAGCGCTGCT--GGCGGTGACTCAAGCGCTAGTGCGAA 1424
QY 752 -GAAC-GCAGG--TGC-----TGGC-----GCCGT-----TCTG--C-CTGGGA 784
DB 1425 TGACTGCGAGGCTGCGCTGTGGCGCGCGCTTACATCCCATCTGAGCGCGCGAGGC 1484
QY 785 CCGG-----GGAACCTCT-----CCTGCGGGAAGCG--GAC--GG- 817
DB 1485 CACGATGAGAGCGCGCTGGGCTTCCCAAGTGTGCGACCAATGCGCTGACAAAGC 1544
QY 818 -CAGGATGG-GCC-----CCAATTCG-----838
DB 1545 CTCAGTACTGAGCGCTGGAAGACAGCTTGGCAGGCAATGCGCTGAAGGAGCGGTGCC 1604
QY 839 -CC-----CTGCGCACTTGACTTC-----ACCAATCCCTTC-----T 871
DB 1605 GCGCGGTGACAGCGCGCGCGCAACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1664
QY 872 GAGAGCTAAACCTG-----TGCTCA-----GGAGCGGAGG-ACTGT 907
DB 1665 TGGGACTGCTGCTGCTGCTGAGCGCGCTCTCACTGAGTGGCGCGCGCGCGCGCGCG 1724
QY 908 G-----AACTTG-----TAGCTGAAGAGCGAG-----AGCTAGCTC 939
DB 1725 GCGACCGAGGTTCCCGACCTCGGCGCTGCGCGA--GGCCAGGCTGTTTCGCAAGAAC 1782
QY 940 TG-----GSCA--CCAGCTGGG-----GAGC-----959
DB 1783 CGACCCCGAGCGCACTGCGTCTGGGCGAGGCGAGCGGGGTGGCGGAGCTGGTGAC 1842
QY 960 -----TCA-----CCCTGCTC-----CCACCCCGCGCG-----982
DB 1843 TCAGAAGGCTCAGGTGCGCTTACCCAGCTTCACTGAGCTTCACTGAGCTTCACTGAGCG 1902

```

QY 983 CAAGTTCTAAGGCTTTTC-----AGAGCGTG----- 1009
Db 1903 CTGTGCTGTGACAGTGTGTGGCCCTGCTGACCCCGAGCGGACACAGAGCGTGTCTCA 1962
QY 1010 -----GAGGTGTG-----GAAGGAGTGTGCTGTCTCCMAAATATGCGAAGGC-----GG 1053
Db 1963 GCAGCCAGGTGTGTGTACATACGGGT--CT-CTCTCCA CGC--CGCCAAAGCCAGCGGG 2017
QY 1054 CGGCAG-----GCTGTCT--TCTGT--CTCTTGGAGAAAGTTCTGTGCGCTG 1102
Db 2018 CGGCGAGCCGCTGGGCGAGCGGCGAGTCTCTCTGATGGAG--CTGCGCGCGCG 2075
QY 1103 ATTATGAATCTA--TAATAGAGTATATAGTTTGT-----TACCTTTTACAGGAAG 1155
Db 2076 CACCCCATCTCACCCCATCATGTTTACAGGTTTCGGCGAGCGGTTTGTTCAGAAAG 2135
QY 1156 GTGACTTCTGTACATGCGATGATA--TTAAACTTTTATATAAGTT 1203
Db 2136 CCGCTCCACCCAGATCGGGTATATAGAGATATGATTTTATTACTT 2186

RESULT 3

US-10-146-731-381
; Sequence 381, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Deanoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary B.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C323
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-146-731-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;
QY 1 CATCTGGGTCTGCCAGAAAGCAATTC--TGGAAGTTAATGTTTGA---GTG 54
Db 191 GCGCGGGGGGAACCGG-----GCGGATTCCTCGCGGTCACCACTGATCCCAT 244
QY 55 ATTTTAAATCTTGTCTGGCG-----GAGAGCGCGCGCTCTCCCGGTATCAGC 103
Db 245 AAACATTATCTC-CGCGGCGCGCGCTCGAGAGCGCGCGCGAGTCCGCGCGCGCC 303
QY 104 GTTCTCTATCTTTGAATCC-GCG-GCTCCGCGGTCTTGGCGTC---AGACAGCGCG 158
Db 304 GC--CCTCGCTGTGCGCCCTGCGCCCTGCG--CACCGCGGCGCGAGCCAGCCAG 359

QY 159 AGAAGCCTCTTTTGAATTTAAGCGGGCTGTG-AAAG-CCAGGGGC-GCGGGGGGGGG 215
Db 360 AGCGGGCGG--AGCG-----GAGCGGGCGAGCTCTGTCGGCGCGCGCGCGCGGG 413
QY 216 GCGAGGCGG---GCCATTTTGAATAAAGA---GGC-----GTGCTTTCCA 255
Db 414 GCGTAGCGGGGGGCGC-----TGATCGGACCCGCGCGGGGAGAGCGGCGCGCGCC 469
QY 256 GGCAGGCTCTATAAGTGAACCGCGCG-----GCGAGCGTGTGCG 292
Db 470 CGAAACGACTTTCAGTCCCGGAGCGCGCGCGCGCAACCCCTAGCATGAAGAGGCGCTCCG 529
QY 293 C-----GC--GTTCCAGTCA-----CTGTAGC---GGACT--TCTTTTGGTT--T 329
Db 530 CTGAGGAGGCGGGCTCTGTG--CATGGTCTGTGGCTCGAGGCTTGGCAGGTGGCAGCC 588
QY 330 TCTTTCTCT-----TTGGG---GCACCTCTGGACTCA-----CT--CCCCAG 366
Db 589 CCAATGCCAGTGTCTGCTATGCTACATAGAGCCCAAGGTGACGACAAGCTGCGCCCGAG 648
QY 367 CA-----TGAGGCGCTGAGCCG--GTGCGGGCTGCTACGAGCGGTGTG--CTGCTG-- 418
Db 649 CAGGCGCTGCGAG--GCTGTGCGCGGTGGGCTATCCCTGTGCGAGCCAGCATCTTCTGTC 706
QY 419 TCGGAA-----CGCA---GTCGT--GCCATGCGCCGG---GCCGAG----- 452
Db 707 ACGCAACCGCATCTGCGCATGTGCGAGTTCGCTGCTGCTGCGCGCAACCTCACCA 766
QY 453 -----GGAAG-----GGCGCG-----GCAGCTG----- 470
Db 767 TCCTGTGGCTGCACTCGAATGTGTGGCGCGAATGTAGTGGGCTGCTTCACTGGCGCTGG 826
QY 471 -----AGGAGCGCTG-----AGC-----TTGC-----TGG----- 491
Db 827 CCTCTGGAGCAGCTGAGCTCAGCATGATGACAGCTCCGCTGTGTGGACCTGCCA 886
QY 492 -----ACGACATGAACCAAC-----TGCTACTCC-----C-GCTTGGGGAAC 527
Db 887 CATTCACGGCTGGGCGGCTACACACGCTGCACTGACCGCTGGACCGCTGCGGCGTGGAGC 946
QY 528 TGGTACCGGAG-----TCC-----TTCG-----C 543
Db 947 TGG--GCCCGGGCTGTTCGCGGCGCTGGCTGCTGCTGCTTACTTACCTGACGACAC 1005
QY 544 GAG-----AGGCACT-----CAGCTT-----AGCCAGG----- 566
Db 1006 GCGCTGAGGCACTGCTGATGACACCTTTCGCGACCTTGGGCAACCTCACACACTTTC 1065
QY 567 -----TGCAAA-----TCTTAC-----AGCGCTCATC-----GACT--ACATCTC 601
Db 1066 CTGACGGAACCGCATCTCCAGGTGCGCGCGCGGCGCTTCCGTGGGCTGACAGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
Db 1126 GACCGTCTCTACTGACACAGAACCGCGTGGCGCCATGTGACCGCGCATGCTTCCGTGAC 1185
QY 618 --TGGCGG-----AGCCAG----- 630
Db 1186 CTGCGCGCTCATGACACTATCTGTTTGGCAACATCTATCAGCGCTGCCACTGAG 1245
QY 631 -CCCTGGACCCCTGATGCGCC--CC--ACCT-----TC-----CCATCC-----A 666
Db 1246 GCGCTGGCGCGCTGCGTGCCTGAGTACCTGAGGCTCAACGACAAACCCCTGGGTGTGT 1305
QY 667 GACAGCGAGCTCG-----CTCGGNACTTGT-----CATCTCAAGACACAA 709
Db 1306 GACTCGCGGCAACCGCCACTCTGGGCTGGGTGAGAAAGTTCGCGGCTCTCTCCGAG 1365
QY 710 AGGAGCTTTTGGC-----ACTGACTCGGCGGTGTCTCT--GACACCT-----CCA- 751
Db 1366 GTGCGCTGAGCTCCCGCAACGCGCTGGCT--GGCGGTGACCTCAACGCTAGTGCACAA 1424
QY 752 -GAAC-GCAGG--TGC-----TGGC-----GCCCGT-----TCTG---C--CTGGA 784

Db 1425 TGACCTGACGGGCTGGCTGGCCACCGGCGCTTACCATCCCATCTGGACCGGAGGC 1484
Qy 785 CCCCG-----GGAACTCT-----CCTGCGGAAACCG--GAC--GG- 817
Db 1485 CACCGATGAGAGCGCTGGGGCTTCCCAAGTGTGCGAGCAGATCGCTGACAAGGC 1544
Qy 818 --CAGGATGG--GCC-----CCAACTTCG----- 838
Db 1545 CTCAGTACTGAGCTGGAACACGAGCTTGGCGAGGCAATGCGCTGAAGGACGCGTGCC 1604
Qy 839 --CC-----CTGCCACTGTGACTTC-----ACCAATCCCTTCC-----T 871
Db 1605 GCCCGGTGACAGCGCGCGGCAACGGCTCTGGCCCAACGCGCACATCAATGACTCACCCCT 1664
Qy 872 GGAGACTAAACCTGG-----TGCTCA-----GGAGCGAAG--ACTGT 907
Db 1665 TGGGACTCTGCTGCTGCTGAGCCCCCGCTCACTGCGAGTGGCGGCCGAGGGCTCCGA 1724
Qy 908 G-----AACTTG-----TAGCTGAGAGCCAG-----AGCTAGCTC 939
Db 1725 GCCACCGGTTCCCGACCTCGGCGCTCGCGGA--GGCAGGCTGTTCAACGAAGAAC 1782
Qy 940 TG-----GCCA--CCAGCTGGC-----GACG----- 959
Db 1783 CGCACCGGACCGACTGCGCTTGGCGCAGGAGCGGGGTGGCGGACTGGTGAC 1842
Qy 960 -----TCA-----CCTGCTC-----CCACCCCAACCCG----- 982
Db 1843 TCAGAGGCTCAGTGCCCTACCCAGCTCAGCTCAGCTCACCCCTCGGCGCTGGCG 1902
Qy 983 CAAGTCTAAGGCTTTTC-----AGAGCGTG----- 1009
Db 1903 CTGGTGTGTGACAGTGTCTTGGCCCTGTGTGACCCCGGACACAGAGGCTGTCTCA 1962
Qy 1010 -----GAGGTCTG-----GAAGAGTGGCTGTCTCCAACTATGCCAAGG-----GG 1053
Db 1963 GCAGCCAGGTGTGTACATAGGGGT--CT-CTCTCAAGC--CGCAAGCAGCGGG 2017
Qy 1054 CGGCAGA-----GCTGGCT--CTGGT--CTCCTTGAGAAAGTTCTGTGGCCTG 1102
Db 2018 CGGCGGACCCGTGGCGGAGGCGAGGCTCTCCCTGATGAGC--CCTGCGCGCCG 2075
Qy 1103 ATTATGACTCTA--TAATAGATATAGTTTTC-----TACCTTTTACAGGAG 1155
Db 2076 CACCCCACTCTCCACCCCATGTTTACAGGTTGCGGCGAGCGTTTGTTCAGAAAG 2135
Qy 1156 GTGACTTTCTGTAAATGCGATGATA--TTAACTTTTATAAAGTT 1203
Db 2136 CCGCTCCACCCAGATCGCGGTATAGAGATATGCAATTTATTTTACTT 2186

RESULT 4
US-10-140-472-381
; Sequence 381, Application US/10140472
; Publication No. US2003013888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Berasini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-140-472-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6, 2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

Qy 1 GATCTGGGTGCTGCCAGGAAAGCAAAATTC--TGGAAGTTAATGTTTTGA-----GTG 54
Db 191 GGCGCGCGGGGAACCGG-----CGGATTCCTCGCGTCAAAACCACTGATCCCAT 244
Qy 55 ATTTTAAATCTTCTGCTGGC-----GAGAGCGCGCTCTCCCGGTATCAGC 103
Db 245 AAACATTTATCTC--CGGCGCGCGCTGCGAGCGCCCGCAGTCCGCGCCGCC 303
Qy 104 GCTTCTCATTTCTTTGAATCC--GCG--GCTCCCGGTCTTCGGCGTC-----AGACAGCCGG 158
Db 304 GC--CCTGCGCTGTGCGCCCTGCGCGCTGCG--CACCGCGCGCCGAGCCAGCCAG 359
Qy 159 AGGAGCTGTGTTGAATTTAAGCGGGCTGTG--AAGC--CCCAGGGCC--GGCGGGGGCGG 215
Db 360 AGCCGGGGG--AGCG-----GAGCGCGCGAGCTCTGCTCCGCGCGCGCGGGCGG 413
Qy 216 GCGGAGGGG--GCCATTTTGAATAAGA--GGC-----GTGCTTTCCA 255
Db 414 GCGGTAGCGGGCGCC-----TGGATGCGGACCGCGCGCGGGGAGCGCGCGCC 469
Qy 256 GCGAGCTCTATAAGTAGACCGCGG-----GCGAGGTGCG 292
Db 470 CGAAACGAGCTTTTCACTCCCGACGCGCGCCGCCCAACCCCTACGATGAAGAGGGCGTCCG 529
Qy 293 C-----GC--GTTGAGGTCA-----CTGTAGC--GGACT--TCTTTTGGTT--T 329
Db 530 CTGAGGAGCGCGCTGTG--CATGGGTGTGTGGCTGCGAGGCTGCGAGGTGGCAGCC 588
Qy 330 TCTTTCTCT--TTGG--GCACCTCTGAGTCA-----CT--CCCCAG 366
Db 589 CCATGCCAGGTGCTCGGTATGTCTAATGAGCCCAAGGTGACGACAAAGCTGCCCCAG 648
Qy 367 CA-----TGAGGCGCTGAGCCG--GTGCGGGCTGTACGAGGCGGTGTG--CTGCTG-- 418
Db 649 CAGGCGCTGCAG--GCTGTGCGGTGGGATCCCTGCTGCCAGCCAGCGCATCTCTG 706
Qy 419 TCGGAA-----CGCA--GTCTG--GCCATGCCCGG--GCCGAG----- 452
Db 707 ACGGCAACCGCATCTCGCATGTGCGAGTGCAGCTTCCGTGCTGCGCAACCTCACCA 766
Qy 453 -----GGAAG-----GGCCG--GCAGCTG----- 470
Db 767 TCTGTGCTGCATCTCGAATGTGTGGCCCAATGTATGCGGCTCCCTTCTACTGCGCTGG 826
Qy 471 -----AGAGCGCTG--AGC-----TTGC-----TGG----- 491
Db 827 CCTCTCTGAGCAGCTGGACCTCAGACGATATGACAGCTCCGGTCTGTGAGCCCTGCCA 886
Qy 492 -----ACGACATGAACCAC-----TGCTACTCC-----C-GCCTGGGGAC 527
Db 887 CATTCCACGGCTGGCGCGCTACACAGCTGCACCTGGACCGCTGCGGCTGCGAGGAGC 946
Qy 528 TGGTACCGGGAG-----TCC----- 543
Db 947 TGG--GCCCGGGGCTGTTCCGGGCGCTGGCTGCGCTGAGTACCTTCTACTGCGAGCAAC 1005

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGTGTCAGGAAAGCAATTC--TGGAAGTTAATGGTTTGA-----GTG 54
Db 191 GCGCGCGGGGAACCG-----GCGAATTCCTCGCGCTCAACCACTGATCCATA 244

QY 55 ATTTTAAATCCCTTGTGGG-----GAGAGGCCCGCTCTTCCCGGTATACG 103
Db 245 AAACAATCACTCTC-CGGCGCGCGCGGTGCGAGCGCCCGCCAGTCCGCGCGCGC 303

QY 104 GCTTCTCATCTTTGAATCC-CGG-GCTCCGGGTCTTCGGCTC---AGACCAAGCGG 158
Db 304 GC--CCTCGCCCTGTGCGCCCTGGCGCCCTGG--CACCGCGCCGAGCCAGCCAG 359

QY 159 AGGAAGCCTTTTGAATTTAAGCGGCTGTG-AAAG-CCCAGGCGC-GCGGCGGCGG 215
Db 360 AGCGGCGCG--AGCG-----GAGCGCGCGAGCTCTGTCGCGCGCGCGCGCGCG 413

QY 216 GCGAGGCGG---GCCATTTGAATAAGA---GGC-----GTGCCCTTCCA 255
Db 414 GCGGTAGCGCGCGCGC---TGATCGGACCCGCGCGCGGCGGAGACGCGCGCGCC 469

QY 256 GGCAGGCTCTATAAGTGACCGCGCG-----GCGAGCGGTGG 292
Db 470 CGAAGACATTCAGTCCCGGAGCGCGCGCCCGCCCAACCCCTACGATGAAGAGGCGTCCG 529

QY 293 C-----GG-GTTCCAGGTCA-----CTGTAGC---GGACT-TCTTTTGGTT--T 329
Db 530 CTGAGGAGGCGCGCTCTGG-CATGGGTCTGTGGCTGCGAGCCCTGGCAGGTGGCAGCC 588

QY 330 TCTTTCCT-----TTGGG--GCACCTCTGGAATCA-----CT--CCCCAG 366
Db 589 CCAATGCCAGGTGCTCGTATGCTCAATGAGCCCAAGGTGACGACAGCTGCCCGCAG 648

QY 367 CA-----TGAAGGCGCTGAGCCCG-GTCCGCGGTGCTACGAGCGGTGTG-CTGGCTG- 418
Db 649 CAGGCGCTGCAG--GCTGTGCGGTGGGCATCCCTGTGTCAGCCAGCGCATTTCTCTG 706

QY 419 TCGGAA-----CGCA--GTCTG--GCCATCGCCGCGG--GCCGAG----- 452
Db 707 ACGGCAACCGCATCTCGATGTGTCAGCTGCGAGCTTCCTGTGCTGCGCGCAACTCACCA 766

QY 453 -----GGAAG-----GCCCG-----GCAGTG----- 470
Db 767 TCCTGTGCTGCACTCGAATGTGTGCGCGGAAATGATGCGGTGCTTCACTGGCGCTGG 826

QY 471 -----AGGAGCGCTG-----AGC-----TTGC-----TGG----- 491
Db 827 CCTCTGGAGCAGCTGGAACCTCAGCGATATGCAAGCTCCGGTCTGTGTGACCTTGCCA 886

QY 492 -----ACGACATGAACCCAC-----TGCTACTCC-----C-GCTTCGGGAAC 527
Db 887 CATTCACGGCTTGGCGCGCTACACAGCTGCACTGGACCGGTGGCGCTGCGAGGAGC 946

QY 528 TGGTACCGGAG--TCC-----C----- 543
Db 947 TGG-GCCCGGGCTGTTCGCGGCGCTGCGGCGCGCTTCCGTGGGCTGCAAGCCTC 1005

QY 544 GAG---AGGCACCT-----CAGCTT---AGCCAGG----- 566
Db 1006 GCGGTGAGGACCTGCTGTATGACACCTTCCGCGACCTGGGCAACCTCACACCTCTTC 1065

QY 567 -----TGAAA-----TCCTAC-----AGCGGTATC-----GACT--ACATCTC 601
Db 1066 CTGCACGGCAACCGCATCTCCAGCTGCGCGAGCGCGCTTCCGTGGGCTGCAAGCCTC 1125

QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
Db 1126 GACCGTCTCTACTGCAACCAAGAACCGCGTGGCGCCCATGTGACCGCGCATGCTTCCGAGC 1185

RESULT 7

US-10-158-790-381

; Sequence 381, Application US/10158790

; Publication No. US20030180879A1

QY 618 --TGGCG-----AGCCAG----- 630
Db 1186 CTTGGCGCCTCANGACACTCTATCTGTGTTTCCCAACAATCTATCAGCGCTGCCACTGAG 1245

QY 631 -CCCTGACACCCCTGATGGCC--CC--ACCT-----TC-----CCATCC-----A 666
Db 1246 GCGCTGCGCCCTCGCTGCGCTGAGTACTGAGGCTCAACGACACCCCTGGGTGTGT 1305

QY 667 GACAGCGAGCTG-----CTCCGAACTTGT-----CATCTCAACGACAA 709
Db 1306 GACTCCGCGGACGCCACTCTGGGCTGCTGAGAAAGTTCCGCGCTCTCTCTCGAG 1365

QY 710 AGGAGCTTTTGGC-----ACTGACTCGGCGGTGTCT--GACACT-----CCA- 751
Db 1366 GTGCCCTGCACTCCCGCAAGCGCTGGCT--GGCGGTGACCTCAAGCGCTAGTGCCTCA 1424

QY 752 -GAC-CGAGG--TGC-----TGGC-----GCGCGT-----CTG--C-CTGGGA 784
Db 1425 TGACTGTGAGGCTGCGCTGTGGCCACCGCCCTTACCATCCCATCTGGAACGCGAGGC 1484

QY 785 CCGCG---GGAACCTCT-----CCTGCCGGAAGCCG--GAC--GG-- 817
Db 1485 CACCGATGAGAGCGCTGGGCTTCCAAAGTGTGCCAGCGAGTCCGCTGACAAGGC 1544

QY 818 -CAGGATG--GCC-----CCAACTTGC----- 838
Db 1545 CTCAGTCTGAGCTGGAAGACCACTTCGCGAGGCAATGCGTGAAGGACGCGTGCC 1604

QY 839 --CC-----CTGCCCACTTGACTTC-----ACCAATCCCTTCC-----T 871
Db 1605 GCCCGTGAAGCGCGCGGCAAGCGCTCTGGCCACGCAATCATGACTCACCCCTT 1664

QY 872 GGAGACTAAACCTGG-----TGCTCA-----GGAGCGAAGG-ATGT 907
Db 1665 TGGAGCTGCTGCTGGCTGTGTGAGCCCGCTCACTGAGTGGCGCGCGAGGCTCCGA 1724

QY 908 G-----AAGTTG-----TAGCTCAAGAGCCAG-----AGTAGTCTC 939
Db 1725 GCCACAGGTTCCCACTCGGCGCTCGCCGSA--GGCCAGCTGTTCGCAAGAAC 1782

QY 940 TG-----GCCA--CCAGCTGGG-----GACG----- 959
Db 1783 CGCACCGCAGCACTGCGCTCTGGCGCAGGCGAGCGGGGTGGCGGACTGCTGAC 1842

QY 960 -----TCA-----CCCTGCTC-----CCACCCCAACCC----- 982
Db 1843 TCAGAGGCTCAGGTGCGCTTACCAGCCTCACTGAGCTCACCCCTGGGCTGGCG 1902

QY 983 CAAAGTTCTAAGGTCTTTTC-----AGAGCTG----- 1009
Db 1903 CTGGTGTGTGACAGTCTTTGGCCCTGTGACCCCGGACACAGAGCGTGTCA 1962

QY 1010 -----GAGGTGTG-----GAAGAGTGGCTGCTCTCCAACTATGCCAAGGC-----GG 1053
Db 1963 GCAGCGAGGTGTGTATACATACGGGT--CT-CTCTCCAGC--CGCAAGCCAGCCGG 2017

QY 1054 CGGCGA-----GCTGGTCT--TCTGGT-CTCTTGGAGAAAGTTCTGTGCCCTG 1102
Db 2018 CGGCGACCCCTGGGCGAGCGCAGCCAGGTCTCTCCCTGATGAGC--CCTGCCGCCGC 2075

QY 1103 ATTTATGACTCTA--TAATAGATATATAGTTTG-----TACCTTTTACAGGAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTGCGCGGCGAGCGTTGTTCAGAACG 2135

QY 1156 GTGACTTCTGTAAACAATGCGATGTATA---TTAACTTTTATAAAGTT 1203
Db 2136 CCGCCTCCCAACCGATCGCGTATATAGATATATGCAATTTTACTTT 2186

[illegible]

```

GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Deanoyers, Luc
; APPLICANT: Flivaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin X
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC448
; CURRENT APPLICATION NUMBER: US/10/158,790
; CURRENT FILING DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
; UN-10-158-790-381

Query Match 47.6%; Score 3593; DB 14; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGGTGCTGCCAGGAAAAGCAAATTC--TGAAAGTTAATGGTTTTGA-----GTG 54
Db 191 GCGCGGGGGGAACCG-----GCGGATTCGCGGCTCAACCACTGATCCCAT 244

QY 55 ATTTTAAATCCTTGCTGGGG-----GAGGGCCGCTCTCCCGGTATCAGC 103
Db 245 AAACATTCATCTC-CGGCGGCGCGGTGCGAGCGCCCGCAGTCGCGCCGCCGCC 303

QY 104 GCTTCCTCATCTTTGAATCC-GCG-GTCCGGGGTCTTCGGGGTC--AGACCAGCCGG 158
Db 304 GC--CTTGCCCTGTGCGCCTTGCGGCCCTGCG--CACCGCGGCCCGAGCCAGCCAG 359

QY 159 AGGAAGCTGTTTGAATTTAAGCGGGGTGTG-AAAG-CCAGGGCC-GGCGGGGGCGGG 215
Db 360 AGCCGGGGCG--AGCG-----GAGAGCGCGAGCTGTCTCCGCGCGCGGCGCGGGCGGG 413

QY 216 GCGGAGGGGG---GCCATTTTGAATAAGA--GGC-----GTGCTTCCA 255
Db 414 GCGGTAGCGGGGGGCC---TGGATGGGACCGGCGCGGGGAGACGGGGCGCCGCC 469

QY 256 GCGAGGCTCTATAAGTGACCCCGCG-----GCGAGCGGTGGC 292
Db 470 CGAAGACGATTTTCAGTCCCGAGCGCCGCCCAACCCCTACGATGAAGAGGGGCTCG 529

QY 293 C-----GC-GTTGCAGTCA-----CTGTAGC---GGACT-TCTTTTGGT--T 329
Db 530 CTGGAGGAGCGCGGTGCTGTG-CAFGGTGTGTGTGCTGAGGCTGGCAGGTGGCAGCC 588

QY 330 TCTTCTCT-----TTGGG---GCACCTCTGGAATCA-----CT-----CCCCAG 366
Db 589 CATGCCCAAGGTGCTGCGTATGCTACAAITAGAGCCCAAGGTGACGACAAAGTGCCTCC 648

QY 367 CA-----TGAAGGCGCTAGCCCG-GTGGCGGGTCTACGAGCGGTGTG-CTGCTTG- 418
Db 649 CAGGGCGCTCAG--GCTGTGCGCGTGGGCATTCCTGTGTGCGAGCGCATCTTCTTCG 706

QY 419 TGGGA-----CGCA-----GTCTG--GCATCGCCCGG---GCCGAG----- 452

```



```
QY 960 -----TCA-----CCCTGCTC-----CCACCCCAACCC-----982
Db 1843 TCAGAAAGCCTAGGTGCTTACCTACCTGAGCTCACCCTCCCTGGGCTGGGG 1902
QY 983 CAAGTCTAAGGTCTTTTC-----AGAGCGTG-----1009
Db 1903 CTGGTGTGTGACAGTCTTTGGCCCTGTGACCCCAAGCCCAAGAGAGCGTGTCTA 1962
QY 1010 -----GAGGTGTG-----GAAGAGTGTGCTCTCTCCAACTATGCCAAGGC-----GG 1053
Db 1963 CGAGCAGGTGTGTACATACGGGT--CT-CTCTCCAGC--CGCCAGCCAGCCGG 2017
QY 1054 CGGAGAG-----GCTGGTCT--TCTGGT-CTCCTTGGAGAAAGTTCTGTGGCTG 1102
Db 2018 CGGCCAGCCCTGGGGCAGCCAGCGGTCTCTCCCTGATGAGC--CCTGCGCCCGC 2075
QY 1103 ATTATGAGCTCTA--TAATAGATATATAGTTTGG-----TACCTTTTATACAGGAAG 1155
Db 2076 CACCCCATCTCCACCCATCATGTTTACAGGTTTGGCGGCAGCGGTTTGTTCAGAGC 2135
QY 1156 GTGACTTTCTGTAACTATGCGATGATA--TTAAACTTTTATAAAGTT 1203
Db 2136 CGGCTCCACCCAGATCGCGGTATATAGATATGCAATTTTATTACTT 2186
```

RESULT 8

```
US-10-137-871-381
; Sequence 381, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Auscin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-381
```

```
Query Match 47.6%; Score 3593; DB 15; Length 2236;
Best Local Similarity 39.8%; Pred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;
```

```
QY 1 GATCTGGGTGTGTCAGGAAAGCAATTC--TGGAAAGTTAATGGTTTGA-----GTG 54
Db 191 GGCCTGGGGGGAACCG-----GGGATTCTCTCGCGGTCAAAACCACTGATCCCAT 244
QY 55 ATTTTAAATCTTGTGGG-----GAGAGGCCCGCTCTCCCGGTATCAGC 103
Db 245 AAACATTCATCTC--CGGGCGCGCGGTGCGAGCGCCCGCCAGTCCGCGCCCGCC 303
```

```
QY 104 GCTTCTCTCATTTCTTGAATCC--GGG--GCTCGCGGTCTTTGGCGTC--AGACAGCGCG 158
Db 304 GC--CCTCGCCTGTGCGCCTTGGCGCCTTGGCG--CACCCCGCGCGAGACCCAGCCAG 359
QY 159 AGGAAGCCTGTTTGAATTTAAGCGGGGTGTG-AAAG--CCACAGGGCC--GGCGGGGGGG 215
Db 360 AGCCGGCGG--AGCG-----GAGCGCGCGAGCCTGTCTCCCGCGCGCGCGGGCGCG 413
QY 216 GCGAGGGGG-----GCCATTTTGAATAAGA--GGC-----GTGCTTCCA 255
Db 414 GCGGTAGCGGGCGGCC-----TGGATGCGGACCGCGCGCGCGGGAGAGCGGGCGCCGCC 469
QY 256 GGCAGGCTCTATAAGTAGACCGCGCG-----GGAGCGGTGCG 292
Db 470 CGAAGAGACTTTCAGTCCCGGAGCGCGCCGCCCAACCCCTACGATGAAGAGGGCGTCCG 529
QY 293 C-----GC--GTTGAGGTCA-----CTGTAGC--GGACT--TCTTTTGGTT--T 329
Db 530 CTGAGGAGAGCGCGTGTCTGG--CATGGGTGTGTGGCTGCGAGGCTGCGAGTGGCAGCC 588
QY 330 TCTTTCTCT-----TTGGG--GCACCTCTGACTCA-----CT--CCCCAG 366
Db 589 CATATGCCAGGTGCTGTGTATGTCTACAAATGAGCCCAAGTGAACAGAAAGTGGCCCCAG 648
QY 367 CA-----TGAAGCGCTGAGCGCG--GTGGCGGGCTGTACGAGGGGGTGTG--CTGCTG- 418
Db 649 CAGGGCTGCGAG--GCTGTGCGCGTGGGCATCCCTGTCTGCCAGCGCGCATCTTCTGTC 706
QY 419 TCGGAA-----CGCA--GTCTG--GCCATCGCCCGG--GCCGAG-----452
Db 707 ACGCAACCGCATCTCGATGTGCGAGTGCAGTGCAGTTCGTGCTGCCGCACTCACCA 766
QY 453 -----GGAAG-----GGCCCG-----GCAGCTG-----470
Db 767 TCTGTGGTGCACCTCGAATGTGTGGCCGCAATTTGATGCGGCTGCTTCACTGCGCTGG 826
QY 471 -----AGAGCGCGTG-----AGC-----TTGC-----TGG-----491
Db 827 CCTCTGAGCAGCTGAGCTGACCTCAGCAGTAATGACAGTCCGGTCTGTGAGCCCTGCCA 886
QY 492 -----ACGACATGAACCAAC-----TGCTACTCC-----C-GCTTGGGGGAAAC 527
Db 887 CATTCACGCGCTGGGCGCGCTACACAGCTGACCTGGAGCGCTGCGCTGCGAGGAGC 946
QY 528 TGTACCGCGAG-----TCC-----TCC-----C 543
Db 947 TGG--GCCCGGGGCTGTTCCGCGCGCTGCGTGGCTGCGAGTACCTCTACCTGAGGACAA 1005
QY 544 GAG-----AGGCACT-----CAGCTT-----AGCCAGG-----566
Db 1006 GCGCTGAGCAGCTGCTGTGATGACACCTTCCGCGACCTGGGCAACCTCACACACCTCTTC 1065
QY 567 -----TGGAAA-----TCCTAC-----AGCGGCTCATC-----GACT--ACATTCTC 601
Db 1066 CTGCACGGCAACCGCATCTCCAGCGTGGCCGAGCGCGCTTCCGTGGGTGACACAGCTTC 1125
QY 602 GAC-----CTGCA-----GCTAGTCC-----617
Db 1126 GACGCTCTCTACTGTGACCAAGACCGCGTGGCCCATGTGACCCCGCATGCTTCCGTGAC 1185
QY 618 --TGGCGG-----AGCCAGC-----630
Db 1186 CTGTGGCGCTCATGACACTCTATCTGTTTGGCAACAATCTATCAGCGTGGCCACTGAG 1245
QY 631 -CCTTGAGCCCTGATGGCC--CC--ACCT-----TC-----CCATCC-----A 666
Db 1246 GCGCTGCGCCCTGCGTGGCTGCGAGTACCTGAGCTCAACGACAAACCTCGGGGTGT 1305
QY 667 GACAGCGAGCTCG-----CTCCGGAACCTGT-----CATCTCAACAGACAA 709
Db 1306 GACTGCGGGCAGCGCCACTCTGGGCTGTGCTGAGAAAGTTCGCGGCTCTCTCCCGAG 1365
QY 710 AGGAGCTTTTGGC-----ACTGACTCGCGCGTGTCT--GACACCT-----CCA- 751
```


QY 528 TGGTACCCGAG-----TCC-----C 543
 Db 947 TGG-GCCCGGGGTGTTCCGCGGCTGGTCCCTGCAGTACCTCTACCTGCAGGACAAC 1005
 QY 544 GAG-----AGCACT-----CAGCTT-----AGCAGG-----566
 Db 1006 GCGCTGCAGGCACTGCTGATGATGACCTTCGCGACTGGGAACTCACAACCTCTTC 1065
 QY 567 -----TGGAAA-----TCCTAC-----AGCGGTGATC-----GACT--ACATTC 601
 Db 1066 CTGCAGGCAACCGCATCTCAGGCTGCCGAGGCGGCTTCGTTGGTGTGACAGCTTC 1125
 QY 602 GAC-----CTGCA-----GGTAGTCC-----617
 Db 1126 GACCGTCTCTACTGCAACGAGACCGGCTGGCCATGTGCACCGCATGCTTCGCTGAC 1185
 QY 618 --TGCGCG-----AGCAGC-----630
 Db 1186 CTGGCGGCTCATGACACTCTATCTGTTGGCAACAATATATCAGGCTGCCCATGAG 1245
 QY 631 -CCTGACCCCTGATGGCC--CC--ACCT-----TC-----CCTATCC-----A 666
 Db 1246 GCGCTGCGCCCTGCGTGGCCCTGCACTACCTGAGGCTCAACGACACCCCTGGGTGTGT 1305
 QY 667 GACAGCGAGTCCG-----CTCGGAACTGT-----CATCTCCACGACAAA 709
 Db 1306 GACTGCGGACACCCCACTCTGGGCTGCTGAGAGTTCCGCGCTCTCTCTCGAG 1365
 QY 710 AGAGCTTTTCC-----ACTCATCGCCGCTGCT--GACACT-----CCA- 751
 Db 1366 GTGCGCTGACGCTCCGCAACGCTGGCT--GGCGGTGACCTCAACGCTAGCTGCCAA 1424
 QY 752 -GAAC-CCAGG--TGC-----TGCG-----GCCGT-----TCGT--C--CTGGGA 784
 Db 1425 TGACTGAGGCTGCGCTGTTGGCCACCGGCTTACCATCCCATCTGACCGGAGGGC 1484
 QY 785 CCGCG-----GGAACCTCT-----CCTGCGGAAAGCG--GAC--GG- 817
 Db 1485 CACCGATGAGAGCGCTGGGGCTTCCCAAGTGTGCGCAGCAGATGCCCTGACAGGC 1544
 QY 818 -CAGGATGG--GCC-----CCAACTTCG-----838
 Db 1545 CTCAGTACTGAGCTGGAACACGACTTTCGGAGGCAATGCGCTGAAGGAGCGCTGCC 1604
 QY 839 --CC-----CTGCCCATTTGACTTC-----ACCAATCCCTTC-----T 871
 Db 1605 GCGCGTGACAGCGCGCGGCAACGGCTCTGGCCCAACGGCAATCAATGACTCACCTT 1664
 QY 872 GGAGCTAAACTGG-----TGCTCA-----GGAGCGAAGG-ACTGT 907
 Db 1665 TGGGACTGCTGCTGCTGTGAGCCCGCTCACTGCAATGCGGCGGCGGCTCCGA 1724
 QY 908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGCTAGCTC 939
 Db 1725 GCCACAGGTTCCCACTCGGGCTCGCGGA--GGCAGGCTGTTCAACGAGAAC 1782
 QY 940 TG-----GCCA--CAAGCTGGC-----GAGC-----959
 Db 1783 CGCACCCGAGCACTGCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1842
 QY 960 -----TCA-----CCCTGCTC-----CCACCCCAACCC-----982
 Db 1843 TCAGAGGCTGAGTGGCTTACCAAGCTCACTGAGCTCACTCCCTCGGCGGCTGGC 1902
 QY 983 CAAGTCTAAGTCTTTTC-----AGAGGTG-----1009
 Db 1903 CTGGTGTGTGACAGTGTGGGCTTGGGCTGCTGAGCCCGGAGGAGGAGGAGGAGGAG 1962
 QY 1010 -----GAGGTG-----GAAGAGTGGCTGCTCTCCAACTATGCAAGGC-----GG 1053
 Db 1963 GCAGCCAGTGTGTATACAGGGGT---CT-CTCTCAAGC---CGCCAGGAGGAGGAGG 2017

QY 1054 CGGAGA-----GTGGTCT--TCTGGT-CTCCTTGAGAAAGTCTCTGTTGCCCTG 1102
 Db 2018 CGGCGGACCCGCGGCGGAGCCAGCCAGGCTCTCTCTGATGGAG--CCTGCGCGCGC 2075
 QY 1103 ATTTATGAACCTA--TAATAGAGTATATAGTTTTG-----TACCTTTTTTACAGGAAG 1155
 Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTCCGCGGAGGTTGTTCCAGAACG 2135
 QY 1156 GTGACTTTCTTATACATGCGATATA--TTRAACTTTTTTATAAAGTT 1203
 Db 2136 CGGCTCCACCCAGATCGCGTATATAGAGATATGCAATTTTACTT 2186

RESULT 10

US-10-141-756-381
 ; Sequence 381, Application US/10141756
 ; Publication No. US20030207359A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P330301C200

CURRENT APPLICATION NUMBER: US/10/141,756

CURRENT FILING DATE: 2002-05-08

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 381

LENGTH: 2236

TYPE: DNA

ORGANISM: Homo Sapien

US-10-141-756-381

Query Match 47.6%; Score 3593; DB 15; Length 2236;

Best Local Similarity 39.8%; Pred. No. 6.2e-171;

Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGGTGTGCCAGAAAGCAAAATTC--TGGAGTTATGTTTGA-----GTG 54

Db 191 GGCGCGCGGGGAAACCGG-----GCGGATTCCTCGCGCTCAACACCACTGATCCCAT 244

QY 55 ATTTTAAATCCTGCTGCGG-----GAGAGCCCGCTCTCCCGGTATCAGC 103

Db 245 AAACATTCATCTC--CGGCGCGCGCGCTGAGGCGCCCGCCAGTCCGCGCGCGCC 303

QY 104 GTTCTCTCATCTTTGAATCC--GGG-GTCCGCGGTCTTTCGGGTGTC---AGACAGCGCG 158

Db 304 GC--CCTCGCCCTGTGCGCCCTGCGCGCCCTGCG--CACCGCGCGCGCGAGCCAGCCAG 359

QY 159 AGGAAGCTGTTTGAATTTAAAGCGGGGTGTG--AACG--CCAGAGGCC--GGCGGGGGCGGG 215

Db 360 ACCCGGGCGG--AGCG-----GAGCGCGCGAGCTCTGTCGCGCGCGCGCGCGCG 413

QY 216 GCCGAGGCGG-----GCCATTTTGAATAAGA---GGC-----GTGCTTCCA 255

Db 414 GCGGTAGCGGCGGCGCC-----TGGATGCGGACCCGCGCGCGGAGACGCGGCGCGCGCC 469

QY 256 GGCAGGCTCTAATAGTACCGCGCG-----GCGAGCGTGGC 292

Db 470 CGAAACGACTTTCAGTCCCGCGAGCGCGCCGCCACCCCTACGATGAGAGGGCGTCGG 529
Qy 293 C-----GC--GTTGAGGTCA-----CTGTAGC-----GGACT--TCCTTTGGTT--T 329
Db 530 CTGAGAGGAGCGCGCTGCTGG--CATGGGTGCTGTGGCTGCGAGCCTTGGCAGTGGAGGCC 588
Qy 330 TCCTTTCTCT-----TTGGG--GCACCTCTGGAATCA-----CT--CCCCAG 366
Db 589 CCAATGCCCAAGTGGCTGCTGATGCTACATAGAGCCCAAGGTGAGCAAGAGTGGCCCGCAG 648
Qy 367 CA-----TGAAGGCGCTGAGCCCG--GTGCGCGGCTGTACAGAGCGGTGTG--CTGCTG-- 418
Db 649 CAGGGCGCTGAG--GCTGTGCGGTGGGCAATCCCTGTGTCAGCCAGCGCATCTTCTGCG 706
Qy 419 TCGGAA-----CCCA--GTCGT--GCCATCGCCCGG--GCCGAG----- 452
Db 707 ACGGCAACCGCATCTCGCATGTGCCAGCTGCCAGCTTCCTGTGCTGCCGCAACTCACCA 766
Qy 453 -----GGAAG-----GCCCG--GCAGTG----- 470
Db 767 TCCTGTGGCTGCACTCGAATGTGTGGCCCGAATTGATGGCGGTGCTTCACTGGGCTGG 826
Qy 471 -----AGGAGCCGCTG-----AGC-----TTGC-----TGG----- 491
Db 827 CCCTCTGGAGCAGCTGCACTGAGCCTAGCGGATTAATGCAAGCTCCGGTCTGTGGACCTTGCCA 886
Qy 492 -----ACGACATGAACAC-----TGCTACTCC-----C-GCCTCGGGAAC 527
Db 887 CATTCACAGCGCTGGCGGCTACACAGCTGCACTGGAGCGCTGGGCTGCGGCTGCAAGGAGC 946
Qy 528 TGGTACCGGAG--TCC-----TCC-----C 543
Db 947 TGG--GCCCGGGCTGTTCCCGGGCGCTGGCTGCGCTGCACTGAGTACCTCTACCTGCGAGCAAC 1005
Qy 544 GAG--AGGCACT-----CAGCTT-----AGCCAGG----- 566
Db 1006 GCGCTGAGGACCTGCGCTGATGACACCTTCGGGACCTGGGCAACCTCAACACCTCTTC 1065
Qy 567 -----TGAAA-----TCTAC-----AGCGGTATC-----GACT--ACATCTC 601
Db 1066 CTGACGCGCAACCGCATCTCCAGGCTGCCGAGCGCGCTTCGGTGGCTGCAAGCGCTC 1125
Qy 602 GAC-----CTGCA-----GGTAGTCC----- 617
Db 1126 GACCGTCTCTACTGACACCAAGACGGGTGGCCCATGTGACCGCGATGCTTCCGTGAC 1185
Qy 618 --TGGCG-------AGCCAGC----- 630
Db 1186 CTGGCGCTCATGACACTCTATCTGTTTCCCAACAATCTATCAGCGCTGCCCACTGAG 1245
Qy 631 -CCCTGGACCCCTGATGGCC--CC-ACCT--TC--CCATCC-----A 666
Db 1246 GCGCTGGCCCGCTGCTGCGCTGAGTACCTGAGGCTCAACGCAACCCCTGGGTGTGT 1305
Qy 667 GACAGCCGAGCTCG-----CTCGGAACCTGT-----CATCTCAACGCAAA 709
Db 1306 GACTCGCGGCAAGCCCACTCTGGGCTGGCTGAGAGTTCGGCGGCTCTCTCCCGAG 1365
Qy 710 AGGACTTTTGGC-----ACTGACTGGCGGTGCT--GACACCT-----CCA- 751
Db 1366 GTGCGCTGAGCTCCCGCAAGCGCTGGCT--GGCGGTGACCTCAAGCGCTAGTGCCAA 1424
Qy 752 -GAAC-GCAGG--TGC-----TGGC-----GCCCGT-----TCTG--C--CTGGGA 784
Db 1425 TGACCTGAGGCTGGCTGGTGGCCACCGGCCCTTACCATCCATCTGGACCGGCGAGGC 1484
Qy 785 CCGCG-----GGAACCTCT-----CTGCGGAAGCG--GAC--GG- 817
Db 1485 CACCGATGAGGAGCGGCTGGGCTTCCCAAGTGTGCGAGCGCATGCGCTGCAAGGC 1544
Qy 818 --CAGGATGG--GCC-----CCAACTTCG----- 838

Db 1545 CTCAGTACTGAGCGCTGGAAGACGAGCTTGGGAGGCAATGCGTGAAGGAGCGGTGCC 1604
Qy 839 --CC-----CTGCCCACTTGACTTC-----ACCAATCCCTTCC-----T 871
Db 1605 GCGCGGTACAGCGCCGCGGCAACGGCTCTGGGCCACGCAATCAATGACTCACCGCTT 1664
Qy 872 GGAGACTAAACTGG-----TGCTCA-----GGAGCGAGG--ACTGT 907
Db 1665 TGGGACTCTGCTGGCTTGTGTGAGCCCGCTCACTGCACTGCGGCGGAGGCTCCGA 1724
Qy 908 G-----AATTG-----TAGCTGAAGAGCCAG-----AGCTAGCTC 939
Db 1725 GCCACCAAGGTTCCCACTCGGCCCTCGCGGA--GGCCAGGCTGTTCACGAAGAAC 1782
Qy 940 TG-----GCCA--CCAGCTGGGC-----GAGC----- 959
Db 1783 CGCACCGCAGCCACTGCGCTCTGGGCCAGGCAAGCGGGGTGGCGGAGCTGGTGAC 1842
Qy 960 -----TCA-----CCCTGCTC-----CCACCCCAACCC----- 982
Db 1843 TCAGAGGCTGAGTGGCTTACCAGCTCACCAGCTCAGCTGAGCTCACCCCCTGGGCTGGCG 1902
Qy 983 CAAGTCTTAAGTCTTTTC-----AGAGGTG----- 1009
Db 1903 CTGGTGTGTGACAGTCTTGGGCGCTGTGACCCCGGAGACACAAGAGCGTGTCA 1962
Qy 1010 -----GAGGTG-----GAAGGAGTGGCTGCTCCAAACTATGCCAAGC-----GG 1053
Db 1963 GCGAGCGAGTGTGTGTAATACGGGT--CT-CTCTCCAGC--CGCCAGCCAGCGCGG 2017
Qy 1054 CGCGAGA-----GCTGGTCT--TCTGGT--CTCCTTGGAGAAAGTCTCTGTGCGCTG 1102
Db 2018 CGGCGGACCGCTGGGCGAGCGCAGCCAGGTCTCTCTGATGGACG--CCTGCGCGCGC 2075
Qy 1103 ATTTATGAACTTA--TAATAGATATAGTTTG-----TACCTTTTACAGGAAG 1155
Db 2076 CACCCCATCTCCACCCCATCATGTTTACAGGTTTGGCGGAGCGTTTTCAGAAAG 2135
Qy 1156 GTGACTTCTGTAAATGCGATGTATA--TTAACTTTTATAAAAGTT 1203
Db 2136 CGCGCTCCCAACCGCATCGGTATAGAGTATGCAATTTTACTT 2186

RESULT 11

US-10-141-759-381
; Sequence 381, Application US/10141759
; Publication No. US20030207361A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P33303R1C197
; CURRENT APPLICATION NUMBER: US/10/141,759
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381


```

RESULT 12
US-10-140-805-381
; Sequence 381, Application US/10140805
; Publication No. US20030207417A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForse, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Collin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C176
; CURRENT APPLICATION NUMBER: US/10/140,805
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-805-381

Query Match 47.6%; Score 3593; DB 15; Length 2236;
Best Local Similarity 39.8%; Fred. No. 6.2e-171;
Matches 808; Conservative 0; Mismatches 360; Indels 863; Gaps 125;

QY 1 GATCTGGGCTGTCCAGGAAAAACAAATTC--TGGAAATTAATGTTTGA----GTG 54
DB 191 GGCGCGGGGGNACCGG-----GCGGATTCCTCGCGCTCAAAACACCTGATCCATA 244
QY 55 ATTTTAAATCTTCTCTGGG-----GAGAGCGCCCTCTCCCGGTATCAGC 103
DB 245 AAACATTCATCTC--CGGCGCGCCGCGCTGAGCGCCCGCCAGTCGCGCGCGCGG 303
QY 104 GTTCTCTATCTTTGAATCC--GCG-GCTCGCGGCTCTTGGCGCTC---AGACGAGCGG 158
DB 304 GC--CTCGCCCTGTGCGCCCTGCGCGCTGCG--CACCGCGCGCGCGAGCCAGCAG 359
QY 159 AGGAAGCCTGTTTGAATTAAGCGGGCTGTG-AAAG--CCGAGGCGC--GGCGGGGGGG 215
DB 360 AGCGGGCGG--AGCG---GAGCGCGCGAGCTCTGTCGCGCGCGCGCGCGCGCGGG 413
QY 216 GCGGAGCGG---GCCATTTGAATAAGA---GGC-----GTGCTTCCA 255
DB 414 GCGGTAGCGCGCGCGG---TGGATGGGACCCCGCGCGCGCGCGCGCGCGCGCGCGG 469
QY 256 GCGAGCTCTATAAGTGACCGCGCG-----GCGAGCTGCG 292
DB 470 CGAAACGACTTTTCAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 529
QY 293 C-----GC--GTTCAGGTCA-----CTGTAGC---GGACT-TCTTTTGTGTT--T 329
DB 530 CTGGAGGAGCGCGGCTGTGCG--CATGGGTGCTGTGCTGAGGCGCTTGGCAGGTGCGAC 588
QY 330 TCTTTCTCT-----TTGGG---GCACCTCTGACTCA-----CT--CCCAG 366
DB 589 CCATGCCAGGTGCTGCTGTATGCTACATGAGCCGAGGTGACGACAGCTGCCCCCAG 648
QY 367 CA-----TGAAAGCGCTGAGCCCG--CTGCGCGGCTGTACGAGCGCGGTGTG--CTGCTG- 418

```

```

DB 649 CAGGCGCTGCGAG--GCTGTGCCCGCTGGGCATCCTGTGTGCCAGCGCGCATCTTCCTGC 706
QY 419 TCGGAA-----CGCA--GTCGT--GCATCGCCCGGG-----GCCGAG----- 452
DB 707 ACGGCAACCGCATCTCGCATGTGCGAGTGCAGCTTCGCTGCGCGCGCAACCTCACC 766
QY 453 -----GGAAG-----GGCCCG-----GCAGCTG----- 470
DB 767 TCCTGTGCTGCACTCGAATGTGCTGGCCCGAATGATGCGGCTGCTTCACTGCGCTGG 826
QY 471 -----AGAGCGGCTG-----AGC-----TTGC-----TGG----- 491
DB 827 CCTCTCTGAGCAGCTGGACCTCAGCATATATGACAGCTCGGCTGTGTGACCCCTGCC 886
QY 492 -----AGACATGAACAC-----TGCTACTCC-----C-GCCTGGGGAAAC 527
DB 887 CATTCACGGCTGCGCGCGCTTACACAGCTGACCTGAGCGCTGCGCGCTGCGAGGAGC 946
QY 528 TGTATCCCGGAG-----TGC----- 543
DB 947 TGG-GCCCGGGGCTGTTCCGGCGCTGCTGCGCGCTGCTGCGCTGAGTACCTCTACTGAGGAGACAAC 1005
QY 544 GAG-----AGGCACT-----CAGCTT-----AGCCAGG----- 566
DB 1006 GCGCTGCGGCACTGCTGATGACAGCTTCGCGACCTGGGCAACCTCAGACACCTCTTC 1065
QY 567 -----TGGAAA-----TCCTAC-----AGCGCTCATC-----GACT--ACATTC 601
DB 1066 CTGCAAGCGCAACCGCATCTCCAGCTGCGCGAGCGCGCTTCCGTGGGTGACACGCTC 1125
QY 602 GAC-----CTGCA-----GGTAGTCC----- 617
DB 1126 GACCGTCTCTACTGACAGAACCGGTGGCCCATGTGACCCGCGATGCTTCCGTGAC 1185
QY 618 --TGGCG-----AGCCAG----- 630
DB 1186 CTGGCGGCTCATGACACTTATCTGTTGGCAACAATCTATCAGCGTGCCTCAGT 1245
QY 631 -CCCTGAGCCCTGATGCGCC--CC--ACCT--TC-----CATCC-----A 666
DB 1246 GCGCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
QY 667 GACAGCGAGCTCG-----CTCCGGAACCTTGT-----CATCTCCAAAGCAAA 709
DB 1306 GACTGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
QY 710 AGGAGCTTTTGGC-----ACTGACTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
DB 1366 GTCCCTGAGCGCTCCCGCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1424
QY 752 -GAAC-GCAGG--TGC--TGGC-----GCGCGT-----TCTG--C--CTGGGA 784
DB 1425 TGACCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1484
QY 785 CCGCG-----GGAACTCT-----CCTGCGGGAAGCGG--GAC--GG-- 817
DB 1485 CACCGATGAGGAGCGCGCTGCGGCTTCCCAAGTGTCTGCGAGCCAGATGCGCTGCAAGGC 1544
QY 818 --CAGGAGTGC--GCC-----CCAACTTCG----- 838
DB 1545 CTAGTACTGAGCTTGGAGACAGCTTGGGAGGCAATGCGCTGAGGGAGCGCGTGC 1604
QY 839 --CC-----CTGCCCACTTGAATTC-----ACCAATCCCTTC-----T 871
DB 1605 GCGCGGTGAGCGCGCGCGCAACCGCTGCTGCGCGCGCAACCGCTGCTGCGCGCGCAACCGCT 1664
QY 872 GGAGACTTAAACCTGG-----TGCTCA-----GGAGCGAAGG--ACTGT 907
DB 1665 TGGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1724
QY 908 G-----AAGTTG-----TAGCTGAAGAGCCAG-----AGCTAGCTC 939
DB 1725 GCCACGAGGTTCCCGACCTCGCGCGCTGCGCGA--GGCCAGGCTGTTTCAACGAAAGAC 1782

```



```

1306 GACTGCCGGGACGCCCACTCTGGGCTGGCTGCAGAAAGTTCCGGCGCTCTCTCTCCGAG 1365
QY 710 AGGAGCTTTTGGC-----ACTGACTCGGCGGTGCT--GACACCT-----CCA- 751
Db 1366 GTGCCCTGCAGCTCCCGCAACGCTGGCT--GGCGGTGACTCAAGCGCTAGCTGCCAA 1424
QY 752 -GAAC-GAAG--TGC-----TGGC-----GCCGCT-----TTG--C--CTGGGA 784
Db 1425 TGACCTGCAGGCTCGCTGTGGCCACCGGCTTACCATCCATCTGGACCGGCGAGGC 1484
QY 785 CCGCG--GGAACCTCT-----CCTCGCGAAGCG--GAC--GG-- 817
Db 1485 CACCGATGAGGCGGCTGGGGCTTCCAAAGTGTGTCACGCAATGCCCTGACAAAGC 1544
QY 818 --CAGGATG-GGC-----CCAACTTCG-----838
Db 1545 CTCAGTACTGAGCTGGAAGACCACTTGGCAGGCAATGCGTGAAGGACGCGTGCC 1604
QY 839 --CC-----CTGCCCACTTGACTTC-----ACCAATCCCTTCC-----T 871
Db 1605 GCCCGGTGACAGCGCGCGGCAACGCTCTGGCCACGCGACATCAATGACTCAACCTT 1664
QY 872 GGAGACTAACTGG-----TGCTCA-----GGAGCGAAGG-AGTGT 907
Db 1665 TGGGACTCTGCTGGCTGCTGAGCCCCGCTCACTGCACTGCGGCGCGGCTCCGA 1724
QY 908 G-----AACTTG-----TAGCTGAAGAGCCAG-----AGTAGCTC 939
Db 1725 GCCACAGGGTTCCCACTCGGCGCTTCCGCGA--GGCAGGCTGTTTACGCAAGAAC 1782
QY 940 TG-----GCCA--CCAGCTGGC-----GAGC-----959
Db 1783 CGACCCGACGCACTGCGGCTTGGCCAGCAGCAGCGGGGTGGCGGACTGGTGAC 1842
QY 960 -----TCA-----CCCTGCTC-----CACCCCAACCCC-----982
Db 1843 TCAGAGGCTCAGTGCCCTACCCAGCTCACCTGACGCTCAACCCCTCGGCGCTGGCG 1902
QY 983 CAAGTTCTAAGGTTTTC-----AGAGCGTG-----1009
Db 1903 CTGGTGTGTGACAGTCTTGGCCCTCTGACCCCGCAGCGGACACAAGAGGTGTCTCA 1962
QY 1010 -----GAGGTGTG-----GAAGGAGTGGTGTCTCTCCAACTATGCCAAGGC-----GG 1053
Db 1963 GCAGCCAGGTGTGTACATACGCGGT--CT-CTCTCCAGC--CGCCAGCCAGCCGCG 2017
QY 1054 CGGCAAG-----GCTGGTCT--TCTGGT-CTGCTTGGAGAAAGTTCTGTGCTGCTG 1102
Db 2018 CGGCCGACCCGTGGGGCAGGCGCAGGCTCTCTGATGAGC--CTGCGCGCCCGC 2075
QY 1103 ATTTATGAACCTA--TAATAGATATATAGTTTTC-----TACCTTTTTCAGGAAG 1155
Db 2076 CACCCCACTCTCCACCCCATCATGTTTACAGGTTTCGGCGGCGAGGTTTGTCCAGAACG 2135
QY 1156 GTGACTTTCTGTAACAATGCGATGTATA--TTAACTTTTATAAAGTT 1203
Db 2136 CGGCTCCACCCAGATCGCGGTATATAGATATGCAATTTATTTTACTT 2186
```

RESULT 14

US-10-142-426-543

; Sequence 543, Application US/10142426

; Publication No. US2004004833A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filwaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C224

; CURRENT APPLICATION NUMBER: US/10/142,426

; CURRENT FILING DATE: 2002-05-09

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 543

; LENGTH: 3721

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-142-426-543

Query Match 47.4%; Score 3576.5; DB 12; Length 3721;

Best Local Similarity 38.9%; Pred. No. 6.2e-170;

Matches 785; Conservative 0; Mismatches 389; Indels 843; Gaps 110;

```

QY 4 CTGGG--GTGCTGCCAGGAAAAAGCAAATTCCTG--GAAGTTAATG--GTTTGTAGTGATT 57
Db 1252 CTGGGCCATGCGCCCTGGATCTCTGGGACCTGATGAGGACCACTGACTGAGTGGCT 1311
QY 58 -----TTTAAATCCTTG-----CT--GGCG 75
Db 1312 GTGGACGTGGGAGCGGCCCTGGGGCAACAGACCGTTGTCTTCTGGTCTTGAGCG 1371
QY 76 GAGA--GGCCC-----GCTC--TC--CCCGGTATCAGCGCTC-----CTC 111
Db 1372 GGGACGGTCTCAAGTCTCTGTCGCGCCCAATGCGACACCTCAGGACGCTGCGGCTC 1431
QY 112 A-----TTC-----TTTGAATCC-----OCGGCTCCGCGT 137
Db 1432 AGTGTCTTCTCTGGAGGAGTTTGAGACCTACCGCGCGACAGTGTGACGCTTCCGGGGCGCT 1491
QY 138 -----CTTCCGGCTCAGACC-----AGCCGGA-----GGAAGCCTGTT 171
Db 1492 GCGAGACAGGCGCGCTGCTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 1551
QY 172 GCAATTT-----AAGCGGG-----CTGT--GAAAGC--CCAG-----199
Db 1552 GCTGCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1611
QY 200 -----GGCGGCGG-----GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 225
Db 1612 TGTATGAAGAACTGTATCGGCACTCAGGACCCCTACTCGGGGTGGGCGGCGGCGGCGGCTCC 1671
QY 226 -----GCCATTTTGAATAAAGAGC--GTGCGCGGCTTCTT-----322
Db 1672 TGCATCTTCTCAGCGCGGCGCACAGACCGGCTTTGAGCAGGAGCTGTCCGGGCGCAGC 1731
QY 254 ----CAGGC-----AGGCTCTATAAGTACCGCGCGCGCGGCGGCGGCGGCGGCGG 294
Db 1732 ACCTCAGGCTTAGGGGACTGCAAGGACTC-CTGCGGCGCAGGCTCTCCGAG-GACCAGC 1789
QY 295 CG-----TTGCAAGTCA-----CTGTAGCGGCTTCTT-----322
Db 1790 CCGGGCTGGTGTGGTGAACCTCTGCTGGTAACTGCTGCTGGTGGCGGCTTCTGCTGGGAG 1849
QY 323 --TTGGTTTTTCT--TTTCTCTTTTGG-----GCACCTCTG-----GACTCCTC---361
Db 1850 CCGTGTGTCCGGCTTCAAGCTGGGCTGTTGCTGGGCTGTTGCTGGGCTGTTGCTGGGCTG 1909
QY 362 CCGAGCATGAAGCGGTGAG-----CCCGGTGCGC--GGCTGCTACGAGGCGGCTG---409
Db 1910 CCGCGCGCAAGACAAAGGAGGCGCATCTCTGCTGCGACGGGCGCG--GCGAGGCGGCTGCTGAG 1967
```


QY 410 -----TGCTCCCTGTCGGAACGAG--TCTGG--CCATCCCGCGGGCCGAGG-----453
Db 1968 CBTACGCCCTCTGGGCGAGCGAGCGCGCAAGGTCCCGGGGGCGGGCGGAGCGGTGG 2027
QY 454 -----GAAAGG-----CCCGG-----CAGCTGAGGAGCGGCTGA-----482
Db 2028 CGGTGGCGCGGGGTTCCCGGAGGCGCTGCT--GGCGCCCTGATGACAGCGGCTGG 2085
QY 483 -----GCT--TGCTGG-----ACGACATGAC-C--ACTGCTACTCCG 516
Db 2086 GCGAAGGCCACGCTGCTGAGGGGGGGGGCCCGACGACCTGAGCTCGGGGCTGCTGCCACG 2145
QY 517 C-----CTGGG--GAA-----GTACCCGAGTCCCGAG--546
Db 2146 CCGAGCAGACGCCGTGCGCGAGAGGCGCTGCGCCACTCGCGACCCCGCCAGCGC 2205
QY 547 --AGCAGTC-AGCTTAG--CCAGGTGGAATCTACAGCGCTCATGACTACTTCT-600
Db 2206 CTGGGGCCCCGGCTTGGGACACGCGCCACCCCTGCT--CCGGCCCTCGGCTTCATCCTC 2264
QY 601 CGACTTGCAGGTAGT-CTTGGCCGAGC-----CAGCC-----CCTG-----GAC 638
Db 2265 CTTCTGCTGCTGGCG 2324
QY 639 CCGCTGATGCGCCCGAC-----CTTCCCATCCAGCAGC-----CGA-----GC 677
Db 2325 CCCC-GACGCGCGCTCTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2383
QY 678 TGCTTCGGAATCTGTCTCTC-----CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Db 2384 TCAACCC-----CAGCG 2440
QY 718 TTGCCACTGCTGCG 774
Db 2441 TGAACCCAGCTCAGCGG--CCGATGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2497
QY 775 -CTGC-----CTGGGACCCCGGGAACCTCTCTGCGGAA-----GCCGGA-814
Db 2498 GCGGCTGAGGAGGCACTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2555
QY 815 -----CGCGAGGATGGGCG-----CCA-----ACTTGGCGCT 842
Db 2556 CCACAGCTTACAGCG 2615
QY 843 GGC---CACTTGAATTT---CACCAATCCCTT-----C-----C 869
Db 2616 GCGGCGCAC-AGACTTGGCG 2674
QY 870 CTGGAGACTAAACCTGGTGC-----C-----TCAGGAG 896
Db 2675 CGTGGCTTAGCG 2734
QY 897 CGAAGGACTGTGAATTTGAGCTTGAAGAGCGCAGCTA-----GCTCTGG--942
Db 2735 CGAGAGACGCTG-CAGAAACGCGG--GGCCCGGGGCAACTCCGAGTGGGTGCTCAAGTC 2791
QY 943 -----CCACGAGC--TGGG-----CGAGCTCAC-----CCTGC 968
Db 2792 CCCCCCGCAGCCACCCCGGAGTGG 2851
QY 969 T--CCACCCCGACCCCGAGTCTTAAGTCTTT-----TCAGAGCTGAGAGGTGGA 1019
Db 2852 TGGCCCTCCCTTACCCCGGGCGCGCAGAGCCTGAGAGGCTTGGGGGTGGGTGGGGCGG 2911
QY 1020 AGGA-----GTGGCTG-----1030
Db 2912 AGGACTTGTGATGATTTGAGTTGACCTTATGCGGTAGTTTGTGTTTTTTTGTGA 2971
QY 1031 -----CTCTCGAA-----ACTATGC-----CAGGCGGGCGG 1057
Db 2972 GTTTTGGTTCTTTTGGGTTTTTCAACCAATGACAACTCCGTTCTCGGGGTGGCGG 3031

QY 1058 AGAGCTG-----GTCTTCTGCTCTCC-TTGGAGAAAG-----TTCTGTTG-----1097
Db 3032 AG-GCAGGGAGGCT--TGGACGCGGTGGGGAATGGGGGCGCACAGCTGCGAGACCTAAG 3088
QY 1098 -----CCCTGATTTTGAATCTCTATAA-----TAGAGTATATAGT---1133
Db 3089 CCCTCCCGCCCGCTGGAAGGTCCTCCCAACCCAGGCGCCCTGCGCTGTGGGTGTG 3148
QY 1134 -----TTGTACCTTTT-----ACAGGAAGT-GAC-----TTTCTGT 1167
Db 3149 CGTGGTGTGCGTCCCTGTTCTGTTGCAAGGGCGCGGGAGTGGCGGTGTGTGCGGT 3208
QY 1168 AACATCGATG--TATATTAACCTTTTATAAAAGT 1202
Db 3209 GCGA--GCGAAGGCTGCTGTGGGCTGTGTCAAGT 3243
RESULT 15
US-10-123-155-543
; Sequence 543, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Betesini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-543
Query Match 47.4%; Score 3576.5; DB 14; Length 3721;
Best Local Similarity 38.9%; Pred. No. 6.2e-170;
Matches 785; Conservative 0; Mismatches 389; Indels 843; Gaps 110;
QY 4 CTGGG--GTGCTGCCAGGAAAGCAATCTG--GAAGTTAATG--GTTTGAAGTAT 57
Db 1252 CTGGGCGCATGCGCTCGATCCTCGGACCTGATGAGCACAGCTGACTGAGTGGCT 1311
QY 58 -----TTTAATCCTG--CT--GGCG 75
Db 1312 GTGACGTTGGAGCGCGCCCTCGGGCAACCGACCGTTGTCTTCTGGGTCTGAGGCG 1371
QY 76 GAGA-GGCGCC-----CGCTC-TC-----CCCGGTATCAGGCTTC-----CTC 111
Db 1372 GGGAGGTCCTCAAGTTCTCTGTCGCGGCCCAATGCCAGCACCTCAGGAGCTCTGGGCTC 1431
QY 112 A-----TTC-----TTTGAATCC-----GGGCTCCCGGT 137
Db 1432 AGTGTCTCTGGAGGAGTTTGAGACTACCGCGCGGACAGGTGTGGACGCGCCCGCGGT 1491
QY 138 -----CTTCGGCGCTCAGAC--AGCCGGA-----GGAAGCTGTGTT 171

1492	GGCGAGACAGGGCAGCGGCTGCTGAGCTTGGAGCTGGACCGCAGCTTCCGGGGGGGGCCTGCTGCT	1551
172	GCAAATT-----AAGCGGG-----CTGT-GRACGC--CCAG-----	199
1552	GCTGCCTTCCCGCGCTGCGTGTGTCCGAGTGCCTGTGGCTCGACAGCAGTACTCGGG	1611
200	-----GGCGCGGG-----GGCGGGGGCGAGCGCG-----	225
1612	TGTAAGAAGAACTGTATCGCAGTCAGGACCCCTACTGCGGGTGGGGCCCCCGACGCGCTCC	1671
226	-----GCCATTTTGAATAAAGAGGC--GTGCCCTTC	253
1672	TGCATCTTCCCTACGCCCGGCACCAAGCGCGCTTTGACAGGACGTGTCCGGGGCAGC	1731
254	-----AGGCTCTATAAGTGACCGCGCGGCGAGCGTGC	294
1732	ACCTCAGGCTTAGGGACTGCACAGGACTC-CTGCGGGCCAGCCTCTCCGAG-GACCGGG	1789
295	CG-----TTGCAGGTCA-----CTGTAGGAGCACTCTT-----	322
1790	CGGGGCTGTGTTCGGTGAACCTCTGTTAACGTCGCGTGGGGCGCTTCGTGGTGGGAG	1849
323	--TTGTGTTTTC--TTTCTCTTTTGGG-----GCACCTCTG-----GACTCACTTC-----	361
1850	CGGTGTGTCCGCTTCAGCGTGGGTGGTTCGTGGGCTCCGTGAGCGCGGGAGCTGG	1909
362	CCACGATGAAGCGCTGAG-----CCCGTGGGC--GGCTGCTACGAGCGCGTG-----	409
1910	CCGGCGCAAGGACAAGAGAGGCCATCTGTGCGCACGGGGCGG--GGCAGGCGGTGTGTGAG	1967
410	-----TGCTGCTGTGCGAAACGACAG--TCTGG--CCATCGCCCGGGCGCGAGG-----	453
1968	CGTCAGCGCGCTGGCGAGCGAGGGCGAGGGTCCCGGGGCGGGGGCGGAGGCGGTG	2027
454	-----GAAGGG-----CCCGG-----CAGCTGAGGAGCGGTGA-----	482
2028	CGGTGGCGCGGGGTTCCTCCGAGGGCCCTGCT--GGCGGCCCTGATGCAGAACGCGTGG	2085
483	-----GCT--TGCTGG-----AGCACAATGAAC--C--ACTGTACTTCCCG	516
2086	GCCAAGGCCACGCTGCTGCAAGGGGGGGCCCCAGACTGAGCTGGGGCTGTGCCCCAG	2145
517	C-----CTGCGG--GAA-----CTG-----GTACCGGAGTCCCAG--	546
2146	CCCGAGCAGACGCGCTGCGCAGAGCGGCTGCCACTCCGACCCCGCACCCCGACGCC	2205
547	--AGGCACTC-AGCTTAG--CCAGGTGGAAATCCTACAGCGCTCATCGACTACATCT-	600
2206	CTGGGCCCCCGCGCTTGAGACAGCGCACCCCTGCT-CCGCGCTCCCGCTTCACTCTC	2264
601	CGACTCGAGTAGT-CCTGGCCGAGC-----CAGCC-----CCTG-----GAC	638
2265	CTCTCTGCTGTGGCGCGCGCCCGGCCCCCGAGCAGCCCCCGCGCTTGGGAGCGCAG	2324
639	CCCTGTATGGCCCCAC-----CTTCCCATCCAGACAGC-----CGA-----GC	677
2325	CCCC-GACGGCGGCTCTATGCTCCGCGCCCGCGCGCTCCACCGGAGCTTCCCGC	2383
678	TGCTCTCGGAACTGTCTATCTC-----CAACACACAAAAGGAGCTT	717
2384	TCACCCC--CCAGCCAGCCCGGACCGCGCGGGTGGTGTCCGCGCCACAGGGCCCCCT	2440
718	TTGCCACTGACTCGGCCCTGTGTGACACCTCCAG-AACGAGAGTGTGGGGCCCGTT--	774
2441	TGGACCCAGCCTCAGCCG--CCGATGGCCCTCCGCGGCGCTGGAGCGCCGCCCGACGG	2497
775	-CTGC-----CTGGGACCCCGGAACTCTCTCTGCCCGAA-----GCCGGA-814	
2498	GCAGCTGAGGAGGCCATGTGGCCCCCAG--CCCTCCGCGCCGACCTGCGCGCAC	2555
815	-----CGGCAAGGATGGGCC-----CCA-----ACTTCGCGCCT	842

Search completed: May 7, 2004, 15:16:57
Job time : 96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:04:14 ; Search time 27 Seconds
(without alignments)
4285.867 Million cell updates/sec

Title: X69111
Perfect score: 7544
Sequence: 1 GATCTGGGGTCTGCCAGGA.....TTAAACTTTTATATAAAGTT 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: Pir1: *
2: Pir2: *
3: Pir3: *
4: Pir4: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1074.5	14.2	1489	2 D70807	hypothetical glyci
2	971.5	12.9	1079	2 F70807	hypothetical glyci
3	971.5	12.9	2639	2 T31328	fibroin - Chinese
4	970.5	12.9	1901	2 F70806	hypothetical glyci
5	930	12.3	2395	1 S50820	surface protein ty
6	898.5	11.9	2718	2 A23475	G surface protein
7	884.5	11.7	1766	2 A42125	trophozoite cystei
8	860.5	11.4	1574	2 T13954	MEGF6 protein - ra
9	860	11.4	1381	2 E70806	hypothetical glyci
10	857	11.4	2704	2 S09118	G surface protein
11	852	11.3	1306	2 A70334	hypothetical glyci
12	800.5	10.6	1660	2 A70869	hypothetical glyci
13	797	10.6	1538	2 H70846	hypothetical glyci
14	789	10.5	1329	2 E70917	hypothetical glyci
15	784.5	10.4	914	2 H70987	hypothetical glyci
16	772	10.2	749	2 A70812	hypothetical glyci
17	739.5	9.8	13288	2 T03399	mucin, submaxillar
18	717	9.5	882	2 E70812	hypothetical glyci
19	714.5	9.5	957	2 D70835	hypothetical glyci
20	706.5	9.4	853	2 A70896	hypothetical glyci
21	687.5	9.1	837	2 E70835	hypothetical glyci
22	684.5	9.1	2352	2 T30201	Notch homolog prot
23	676.5	9.0	2531	2 S18188	notch protein homo
24	675	8.9	1620	2 T27283	hypothetical prote
25	675	8.9	2555	2 A40043	notch protein homo
26	670.5	8.9	5376	2 T42215	zonadhesin - mouse
27	669.5	8.9	2533	2 T28674	alpha-5ID-immobili
28	668	8.9	778	2 F70963	hypothetical glyci
29	666	8.8	481	2 A35628	loricrin - mouse

ALIGNMENTS

RESULT 1

D70807
hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807
R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: D70807
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1489 <COL>
A: Cross-references: GB: AL02022; GB: AL123456; NID: g3261554; PIDN: CAAL7751.1; PID: g2924
A: Experimental source: strain H37RV
C: Genetix:
A: Gene: RV3514
C: Superfamily: collagen alpha 1(IV) chain

Query Match	14.2%	Score	1074.5	DB 2	Length	1489			
Best Local Similarity	33.7%	Pred. No.	3.5e-42						
Mismatches	432	Conservative	19	Mismatches	561	Indels	271	Gaps	54
QY	1	GATCTGGGTGCTGCCAGGAAAAAGCAAAATCTCGA--AGTTAATGTTTTG--AGTGAT	56						
Db	208	GGLIWGGGGAGVGAGGGTGGAGRAELLFAGAGGAGACTDGGPGATGCTGGHGVGGD	267						
QY	57	TTTTAAATCTTGTCTGGCGAGAGGCCCGCTCTCCCGGTATCAGCGCTTCTCTCATCTCT	116						
Db	268	GGWLPAGGAGGAGGGGAGGAGSDG-----GALGGTGTGG-----	303						
QY	117	TTGAATCCGGGCTCCCGGCTCTTCGGCGTCAGACCCAGCGGAGGAGCCTGTTTGAAT	176						
Db	304	TGAGAGAGRGALLLAGGQ-----GSLGGAGGQ-----GTTGGVGGT	350						
QY	177	TTAAGCGG--GCTGTGAACCCAGGCGCGGGCGGGCGGAGCGGCGCCATTTTG	234						
Db	351	GGKGGVGVAGLGGAGGAAGQLFSAAGNAGVGGAGGGGGGAGGAGADADQPGATG	410						
QY	235	AATAAGAGGCGTGCCTTCAGGCGAGGCTCTATAAGTGACCGCGCGCGGAGCGGTGCG--	292						
Db	411	GTGFAGGAGGAGG-----AGGSSG-----AGTTNGSGAGGCGGAGGAGGADNP	456						
QY	293	CGCGTTGCAGGTCACTGTAGCGGACTTTTGTGTTTTTCTTTCTTTTGGGGCACCTCTG	352						
Db	457	TGIGTGGDGTGGAAGAGGAGGA-----AGTGGTGMIGTTGNAGVGGAGGGGGGAG	511						
QY	353	GACTCACTCCCGCA-TGAAGGCGCTGAGCCCGGT-GCGCGGCTGTCTACGAGCGG--	407						

notch-1 protein -
notch3 protein - h
Balbian ring 3 pr
alpha-5ID immobili
hypothetical glyci
notch 3 protein -
hypothetical glyci
hypothetical glyci
notch homolog se
hypothetical glyci
major ampullate fi
surface antigen - p
transmembrane prot
Xotch protein - Af
period clock prote
surface protein 51

30 661.5 8.8 2531 2 A46019
31 656 8.7 2321 2 S78549
32 654.5 8.7 1700 2 S08167
33 644.5 8.5 2533 2 T28675
34 638.5 8.5 801 2 F70824
35 637.5 8.5 2318 2 S45306
36 635.5 8.4 783 2 E70824
37 634.5 8.4 767 2 E70895
38 631 8.4 2531 2 T31070
39 625 8.3 731 2 C70974
40 618.5 8.2 718 2 A30668
41 618.5 8.2 2543 2 T31687
42 616.5 8.2 2437 2 S42612
43 613.5 8.1 2524 2 A35844
44 603.5 8.0 713 1 UMMS
45 601.5 8.0 2233 2 T28669

RESULT 2
 B70807
 hypothetical glycine-rich protein RV3512 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: B70807
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Geniales, S.; Hamlin, N.; Holroyd, S.
 ; Jones, R. W. A.; Bagers, J.; Butler, S.; Seeger, K.; Skelton, S.; Squares, S.

QY 775 CTGCTGGGACCCGGGAACCTCTCTGCGGAAGCCGACGCGAGGATGGCCCAAC 834
Db 760 GTGGGGTGGDGGNGGNGDNTAN-MTAQAGSDGG--NGDGGFGGAGAGGG-----809
QY 835 TTGCGCCCTGCCACTTGACTTACCAATCCCTTCTCTGGAGACTAAACCT-----884
Db 810 -----GLTAGANGTGGGGAGG-----DGGNGAIGGHHPLTDDPGNGGTG 850
QY 885 --GGTGTCTAGGAGCGAAGACTGTGA---ACTTGTAGCTGAAGAGCAGAGTAGCT 938
Db 851 GNGGTGGT--GGAGIGSLGGGTGGDGGNGGNGGTGGGEGVGGAGGTGAAGNGSDG--906
QY 939 CTGCGCCACGAGTGGGCGAGCTCACCTGCTGCCACCCACCCCAAGTCTTAAGTCTT 998
Db 907 TGG-----TGGSDG-----GAGTGTGT 923
QY 999 TTCAGAG---CGTGGAGGT-GTGGAGAGAGTGGCTGTCTCTCCAAACTATGCCAAGCGG 1053
Db 924 GGTGGLDPRVGGSDGGTGGGGAAGNGGNGAG-----AGNGNGG 968
QY 1054 CGCAGAGCTGGTCTTCTGCTCTCTTGGAGAAAGTCTTGTGCCCTGATTATGAAC 1113
Db 969 TGGAGGIGGTG-----NGDAPGVPYPGAGGAGGTGGKGG---TGGNGSGTG---1016
QY 1114 CTATAATAGAGTATATAGGTTTGTACCTTTTTCACGGA-----AGGTGACTTTCTG 1166
Db 1017 -----SGGTGGDGTGGGGNGGTGWNKGDTGSGGAGDGGKAPAGGTG-----1062
QY 1167 TAACAATCGGATGA 1181
Db 1063 -----GAGDGGGA 1070

RESULT 3
T31328
fibroin - Chinese oak silkworm
C/Species: Antheraea pernyi (Chinese oak silkworm)
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C/Accession: T31328
R/Seizutsu, H.; Tamura, T.; Yukuhiro, K.
A:Submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, Anth
A:Reference number: 220995
A/Accession: T31328
A/Status: preliminary; translated from GB/EMBL/DBJ.
A/Molecule type: DNA
A/Residues: 1-2639 <8EZ>
A/Cross-references: EMBL:AF083334; NID:G3450882; PID:G3450883; PIDN:AAAC32606.1
C/Genetics:
A/Introns: 14/3

Query Match 12.9%; Score 971.5; DB 2; Length 2639;
Best Local Similarity 25.1%; Pred. No. 2.2e-37;
Matches 386; Conservative 60; Mismatches 707; Indels 385; Gaps 43;
QY 2 ATCTGGGGTGTCTGCAGGAAACCAATT-----CTGGAAGTTAATGTTTGA- 51
Db 344 AASGAGRGDGGYGGSSSNAIAAIAAARRAGHRAAGSNAIAAIAAASGAG 403
QY 52 -----GTGATTTTAAATCTTGTGCGCGAGAG-----GCCCGC 86
Db 404 GSGGGYGMWDGGYSDSNAIAAIAAAGSAGAGGGYGMWDGGYSDSNAIAA 463
QY 87 CTCCTCCCGGTATCAGCGCTTCTCATTTTGA-ATCCCGGCTCCCGGTCTTCGGCG 145
Db 464 AIAAASGAGGGGGYGGYSDSNAIAAIAAAGAGAGGAGGAGGGYGMWDGGYSDSNAIA 523
QY 146 TCAGACCGCGGAGGAGCTGTTTGAATTTAAG-----GGGTGTGAACGCC 197
Db 524 AIAAIAAAGSAGGGGDDGGYGGSSSNAIAAIAAARRAGHDSAAAGSNAIAAIAA 583
QY 198 AGGCGCGCGGGCGGGCG-----CGAGGCGG-----C 227

Db 584 AASG-AGSGGGYGMWDGGYSDSNAIAAIAAAGSAGAGGGYGMWDGGYSDS 642
QY 228 CATTTTGAATAAAGAGCGTCCCTTCCAGGAGAGCTCTAATAGTGAACGCCCGCGCAGC 287
Db 643 AIAAIAAIAAASGARGSGG-----YGGYSDSNAIAAIAAIAAAG-SGAGGCGGY 694
QY 288 GTGCGCGGTTGCGAGTCACTGTAGCG-----GACTTCT-----321
Db 695 GWDG-GYGSNAIAAIAAIAAAGSAGAGGGYGSNAIAAIAAIAAARRAGHD 753
QY 322 TTTGTTTCTTCTTCTTGGGCACTCTGGAATCTATCCCGAGCATGAAGGCGCTGAG 381
Db 754 SAASNAIAAIAAIAAASGAGSGGGYGMWDGGYSDSNAIAAIAAIAAAGSAGAG 813
QY 382 CCCGCTGCGCG-----CTGCTACGAGCGGTGTGCTGTCTCGGAACGAG 429
Db 814 ---GGYGMWDGGYSDSNAIAAIAAIAAASGARGSGGYGSNAIAAIAAIAAAG 870
QY 430 TCTGCCATCCCGCGGCGAGGAGGCGCGAGCTGAGAGCGCGCTGAGCTTGTCT 489
Db 871 SGAGGCGGYGMWDGGYSDSNAIAAIAAIAAAGSAGGAGGCGG-----S 917
QY 490 GGACGACATGAACCA-----CTGCTACTCCCGCTCGGGAACCTGTTACCGGAGTC 541
Db 918 GSSNAIAAIAAIAAARRAGHDSNAIAAIAAIAAAGSAGSGGYGMWDGGYSD 977
QY 542 CCGAGAGCACTAGCTTTAGCCAG---GTGG-----AAATCTTACAGC-CCGTCTAT 588
Db 978 SAAIAAIAAIAAAGSAGGAGGGYGMWDGGYSDSNAIAAIAAIAAASGARGSGGYG 1037
QY 589 CGACTACATTTCTGACCTGC-AGGTAGTCTCTG---GCCGAGCGAGCCCTTGGACCCCT 643
Db 1038 YGSDSNAIAAIAAIAAAGSAGGAGGGYGMWDGGYSDSNAIAAIAAIAAAG-----S 1091
QY 644 GATG-----GCCCGCACCTTC-----CCATCCAGACAGCCGAGCTCGC 681
Db 1092 GAGRGDGGYGSNAIAAIAAIAAARRAGHRAAGSNAIAAIAAIAAASGAGSGGG 1151
QY 682 TCCGGAATCTGTCTCTCAACGACAAAGAG-----CTTTTGCACCTGAC 728
Db 1152 YGMWDGGYSDSNAIAAIAAIAAAGSAGSGGYGSNAIAAIAAIAAAGSAG 1211
QY 729 TCGGCGGTGTCTCTG---ACACTCTCAGAACAGTGTGCGCGCTTCTGCTCGGAC 785
Db 1212 GAGGCGYGMWDGGYSDSNAIAAIAAIAAASGAGSGGYGSY-GYSDSNAIAAIAA 1270
QY 786 CCCGGAACCTCTCTCTGCGG-----AAGCCGAGCGGAGGATGG-GCC 829
Db 1271 AAGGSGAG-----GAGGGYGMWDGGYSDSNAIAAIAAIAAAGSAGGAGG 1322
QY 830 CCAACTTCGCGCTGCCACTTGTACTTCAACCAATCCCTTCTCTGGAG-----875
Db 1323 SNAIAAIAAIAAARRAGHRAAGSNAIAAIAAIAAASGAGSGGYGMWDGGYSDS 1382
QY 876 -----ACTAAACCTGGT-----CTCAGGAGCGAAG-----902
Db 1383 AIAAIAAIAAIAAAGSAGSGGYGSNAIAAIAAIAAAGSAGGAGGGYGMWDGG 1442
QY 903 -ACTGTGACTGTGCTCTGAGCCAGAGCTAGCTTCTGCCACCGAGCTGCGCGAGCTC 961
Db 1443 YGSDSNAIAAIAAIAAAGSAGSGGYGSNAIAAIAAIAAASGAGGAGGYG--1500
QY 962 ACCCTGTCTCCACCCCG 1004
Db 1501 -----GYGSDSNAIAAIAAIAAAGSAGGAGGGYGMWDGGYGSNAIAAIAAIAA 1555
QY 1005 GCGTGGAGGTGGA-----AGGAGTG 1026
Db 1556 GAGRGDGGYGSNAIAAIAAIAAARRAGHRAAGSNAIAAIAAIAAASGAGGAGGG 1615
QY 1027 ---GCTGTCTCCAACTATGCCAAG-GCGGCGGAGAGCTGGTCTTCTGCTCTCTTGG 1082

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
 C;Genetics:
 A;Genetic code: SGC5
 A;Introns: 472/3; 1310/3; 1821/3
 C;Superfamily: G surface protein

Query Match 12.3%; Score 930; DB 1; Length 2395;
 Best Local Similarity 23.4%; Pred. No. 1.6e-35;
 Matches 481; Conservative 67; Mismatches 637; Indels 874; Gaps 111;

QY	3	TCCTGGGTG-CTGCAGGAAACCAAT-----TCTGGAAGTAAATGGTTTGGCT	53
DB	248	TCKSFRDGTCTTKANGCGVTRTCAATTOAACVKNSSNGDCYWTGTACVCKTCNA--	305
QY	54	GAITTTTAAATCCCTTGTCT-----GGCGGAGA-----GGCGCGGCTC--TC	91
DB	306	PATWTNSACAGFTVGTCTKSGGCVNAGCAANVQAACVKNSTGDCIWDITCKEKC	365
QY	92	C-----CGGTATC-----AGCGTCTCTCATCTTTGA-----ATCC--GC	126
DB	366	ANAPTTNNHDLCTSYLSCTVKSGGCQPRCTCANAPTITINDACEAYLPANNCLTKSG	425
QY	127	GGC-----TCC-----GGGTCTCTGGGTCTCAGACCGCGAGGAACCTGTTT	171
DB	426	GGCVPTTCAVITLACVKNSSGATCFWDSASSTCKDKTCAN-APSTNTNLDLCAFLS	484
QY	172	GCAATTTAAGC-----GGGC-----TCTGAAC--193	
DB	485	SCTVNSTNAGCVDKTCNSLAQITCDKDLNNKACIMWKGKYKRECVLASSTTATHADCQT	544
QY	194	--GCCCA--GGGC-----CGG-----CG-GGGCGGGGC--217	
DB	545	YDVGCTLSNTGCVPLPLKCEALITBIACNIRLQVTSVKVSYPLGWNSSCIDKACST	604
QY	218	--CGA--GGCGG-----GC-----CATTTG-----234	
DB	605	APKTTATTSDCGTYSKSGVANNPVNGSIQCDLPTTCAARKSTENCEIIRTFPTCLMN	664
QY	235	AATAAGAGGGTCCCTT-----CCAGGCAGGCTCTAAGTAGACCGCGCGG	282
DB	665	SATSACVEKSCSTASVTTTGTFLVPSNTWCLAYLSNACIANNTADGCIPEKSSCGSLT	724
QY	283	CGAGCGTGC-----CGCGGTGACGGTC--ACTGTACGGACTCTTTTGGTTTCTTTC	335
DB	725	TSANGCVGSKANGCYWNGTTCVDRCTSNISLSHSCNGVLTNC--TVNNAKTIQSLA	782
QY	336	TCTTTGGGCAC-----CTCTGGAC--TC-----ACTC-----CCCA	365
DB	783	TACTSYGSENCKQTSAGKTCIWTGTACRNATCLDASDTNSFSDGECSEAYTPSDCTV	842
QY	366	GCATGAAGCGCTGAGC-----CCGGTCCGGCGGTGCTACG-----AGCG--406	
DB	843	LYKTGAV-GCVTRSANCSYVQAQVRLTLTAASDDCTWKSSSHKCFPSNTYLSGACSTFQ	901
QY	407	GTGTGC-----TGC-----CTGTGC--GAA--CGCAG--TCT--GGCCAT	438
DB	902	GTKLCEAIKTCNNAVAVETDTCFCAVVTGTSIETHAGCQYSTTCTANSGGTACLT	961
QY	439	CGCCC-----GGGGC-----CGAGGGAAGGC-----CGGCGACTGAGGAGCGCT	480
DB	962	LAAACSTVTAQGDVRRDGLSCFWNAGASPAAQDITSTNCNALSITGATHASCOAYS	1021
QY	481	GAGCTTGTGGAC-----GACATGAACCC-----TGTACTCCG--CTTGGG	523
DB	1022	TA-CTSI SDGTACQAIQAACSSYSALTACTKRTDGLKCFWNTGATPAAACQDITSTNCNA	1080
QY	524	GAACGTGTACCGGAGTCCCGAGAG-GCACTCAGCTT-----AGC-----CAGGTGGA	570
DB	1081	LASLTGATHATCAQVSTACTSISDGTACQAIQAACSSYSALTACTQRTDGLKCFWNTGAT	1140
QY	571	AATCCTACAGCGCG-----TCATCGACTACATTC-----CGACCTGAGGTAGTCCT	618
DB	1141	PAAACQDITSANCNALSITGATHASCOAYSACTSISDGTACQALQAACSSYSQAQACT	1200

QY	619	GGCGGAGC-----CAGCCCT-----GGAC--638	
DB	1201	QRTDLKCFWNTAATPAAACQDITSAACSLITGLSSATHATCAQYSTACYLNIAGNACQA	1260
QY	639	---CCCTGATGGCCGCCACCTTCC-----CAT---CCAG-----ACAGCC--673	
DB	1261	FSTCEALTGSLNLTWITICQAFSTTCSVKRDGTGCVTIQASACTGYTTIANCYSTAGYCTAN	1320
QY	674	-----GAGCTC-----GCTCCGGAACCTTGTCTCATCT-----698	
DB	1321	SGDSACQAIASSTICEQIKLGSSPAFDDTKNTFKTGICIALGTSGCQTKTCAKNTTPPAH	1380
QY	699	---CCACGACA---AAAGGAGCTTTTGCACGTGCTC-----GGCGGTGCTC-----741	
DB	1381	SDCNTWLSTCSNNAVSSPTACTTMAATCASLTSTCVVAVEGECVWGSVCVRKTCDTAS	1440
QY	742	-----GACAC--CTCCAGAACGCGAGTGC-----TGG-----766	
DB	1441	ADTSFDSHAEICIAVLSTCTVARTGCCQARATCASYSKSSQCKFNSTGKCFWNPNNKTCV	1500
QY	767	---CGCCCGTT-----CTGCTGGGACCCCGGGAACCTCT-----CCTGC	803
DB	1501	DLNCGNIEATTYDTHAECAVADTQLLCTVRATNGAAVPGCMARGACSSYSIEQCCKTNP	1560
QY	804	CGGAA-----GCGGACCGCGAGGATG-----GCGGACCGCGAGGATG	825
DB	1561	SGGVCMVNTLTPVCQDKSCTTAPTATATHADCDSYFSTATIKTVVATPTDNGGAVAL	1620
QY	826	GGCCCAAACTT-----CG-----CCCTGC-----CCACTTG	851
DB	1621	GGCQQTAAACSSYIHOEQCRFNATGDLGWNGTQCADSKSCATAPATDYDDNDKCRAYFNN	1680
QY	852	ACT-----TC-----ACCAATCCCT--TCCTGGAGACTAAAC	882
DB	1681	KTVASSGGGCVDPDTCESMTQKCVSDKTRSCYVWNGTACITFTCENAPDSTASAEBC	1740
QY	883	CTGTGTCT-----C-----AGGAGGAGGACTGTGAACTT--GT-----AGCC	919
DB	1741	NTYLAGCTLDSVKCKTKVCEDFAFATDALCKSALSTCTTNGTNCVTRGTCTCFQAQNAQCV	1800
QY	920	TGAAGAGC-----CA-----GAGCTAGTCT	939
DB	1801	TSSANVQCEWMPAVGTNOAYCTVKTONTAPATLISESACASYFNCITKNGGCGVTKSTC	1860
QY	940	T-----GGCC-----ACC-----AGCTGGCGGAGC--959	
DB	1861	SAVTIDVACTTALNGTVCAWDSAQNKCRDKCQDFSGTTHAACQTRAGCTAGANGKCAR	1920
QY	960	--TC-----ACCTGTCTCC-----ACCCACCCCA--AGTCTTAAGT	995
DB	1921	VQNCBQTTIRISACTEBGTNGPCLWNNYVNSDGSKGACPRYTSCKSLAWSNDSCKWISNQ	1980
QY	996	CTTTTCAGAGC-GTGGAGGTGTGAAGGAGTGGC-----TGTCT--1032	
DB	1981	CTTN---GSNCIGITLCTSETNTDGGCVSGYDGCISQVPALENSDPKVKPVTSCADAFY	2037
QY	1033	-----CTCCAACT--ATGCCAAGCGGC--GGC-----AGAGCTGTCTTCTGTCTC	1077
DB	2038	TTTSDCQTASKCKTNGTGTGICALGACSSYTTQAGCYFNDKGALTSGLTST--GICTW	2095
QY	1078	CTTGAGAAAAGGTCTGTGTGCCCTG--ATTATGAAC--TCTA-----TAATAGAGT--1125	
DB	2096	DTTASCRDQSCADLTGTTHATCCSQLSTCSDGTSCLVKGACTSYTTTCTACTAVGSDG	2155
QY	1126	-----ATAT-----AGGTTT-----TGTAC-----CTT-TTT	1146
DB	2156	ICYWELASSSTNNNTAKRLLACADITQNGTSTNVCAVALSSCVSNSTACIAKNCSTYTTK	2215
QY	1147	TAC-----AGGAAGGTG-----ACTTCTGTAAACA-----TGGGATGT	1180
DB	2216	TACNSGLDGCIVFTQSTATGAAGTGTCACTTANSQVACQAAKDRCSWTAASGT	2275

Qy 1181 ATATTAAACTTTTATATAA 1199
Db 2276 GTTAVASKCAHTTCATNOA 2294

RESULT 6
G surface protein - Paramesicium primaurelia
C:Species: Paramesicium primaurelia
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 07-Dec-1999
C:Accession: A23475
R:Prat, A.; Katinka, M.; Caron, F.; Meyer, B.
J. Mol. Biol. 189, 47-60, 1986
A:Title: Nucleotide sequence of the Paramesicium primaurelia G surface protein. A huge pro
A:Reference number: A23475; MUID: 87060934; PMID: 3783679
A:Accession: A23475
A:Molecule type: DNA
A:Residues: 1-2718 <PRA>
A:Note: the authors translated the codon TGC for residue 2665 as Trp
C:Genetics:
A:Genetic code: SGC5
C:Superfamily: G surface protein

Query Match 11.9%; Score 898.5; DB 2; Length 2718;
Best Local Similarity 24.68; Pred. No. 4.4e-34;
Matches 453; Conservative 56; Mismatches 630; Indels 701; Gaps 102;

Qy 3 TCTGGGGGTGTCGACGAAAAGCAATCTGGAAGTTAATGGT
Db 835 TCTVYKVGAGCC-----VXSANCSVMYSAQCHRTLTNLNLTANDCKMIVDRCYALSSFA 890
Qy 53 TGATTTTAAATCC---TTGCTGGCGAGAGGCCCTCTCCCGGTACAGCGCT--T 107
Db 891 TGACTTTPKTKMCEGYRACNTVGAASASCTLDCTLKT-----GSLITAD 939
Qy 108 CCTCAATCTTTGAATCCGCGGTCCTCGCGCTTCGCGGTCAGACGAGGAGGAGCCT 167
Db 940 CQALDSTCSVKKGGT-----GCIATQSTCAGYSTAANCPR---SSASGTAGYC- 985
Qy 168 GTTTGCAATTAAGCG--GGCTGTG-----AAGCCGACAGGC-----CGCGGGGG 211
Db 906 AMNTNCQSVSAECAFVGLTGLDHSKCOLYHSSCTSLKDTGCGQBYKTCGVAATNN 1045
Qy 212 C-----GGGCGCAGGC--GGGCCATTTTGAATAAGA--GGC-----GTCC--CTTCCAG 256
Db 1046 CATSGQKCFDVECLRFNSCASITGTLTATCGTYDAGCVANVNGTACQEKLATCDLY 1105
Qy 257 GCAGCTCTATAAGTGACC-----GCCGCGGAGCGTGGCGGTTCAGGTCACCT--- 308
Db 1106 LTQNSCSTSAATADKCAWSGTACLAATVTVGTHCAVYVGTGLTDLICAAYNANTANKA 1165
Qy 309 GTAGCGGACTCTTTGGTGTCTTTCTCTTTGGGCGACCTCTGGAC-----TCACCTCCC 364
Db 1166 GTA--CQEKATCNLYT--TEATCSTSAATAADK-----CAWSGAACLAATVTVATECAVY 1217
Qy 365 AGCATG-----AAGGCGCTG--AGCCCGGTGGCGGCTGCTACGAGCGGTGCTGCC 416
Db 1218 TG--TGLTDLICAAYNANTANKAGTACQEKATCNLYTEATCSTSAATADK----- 1271
Qy 417 TGTGGAAACGAGTCTGGCCATCGCCCGGCGGAGGAGGCGGCGGCGGCTGAGGAGC 476
Db 1272 --AWSGAACLAATVTVATECAVYVGTGLTNAIC-----AAYNANTANKAGTACQEKAT 1323
Qy 477 CG--CTGAGCTGTGTCGACGACA--TGAACCA-----C-----TGCTACTCCC--GGCT 519
Db 1324 CNLYTTEATCSTSAATADKCAWSGAACLAATVTVATECAVYVGTGLTNAICAAYNANT 1383
Qy 520 CGGGAACCTGTAC-----CGAGTCCGAGGAGGACCTCAGC-----TTAGCCA--G 565
Db 1384 ANKAGTACQEKATCNLYTTEATC--STSAATADKCAWSGAACLAATVTVATECAVY 1440
Qy 566 GTGGAAATCCCTACAGC-----GCCTCATCGACTACATT----- 598

Db 1441 GTGLTFAICATYNAGCINIKDGTGCBQBAKANKDYTTSNKTAQTITSLSLWIDNSCYP 1500
Qy 599 -----CT-----CGACCTCAGGTAGTCCTG--GCCG--AGCCAGC-----C 631
Db 1501 VTDLNCNVITGLGFVHAQCOAVSTGCTSVSDGSKQDFKSTCEQVPGTTLGCTKTASTKC 1560
Qy 632 CCTGGAC-----CCCTGATG-----GC-----CC-----C--CA 654
Db 1561 YLGSSACIISNVATDCAKITGSAGTIFFECQSVNTGCSVNRARSACVQQAQCSGYTS 1620
Qy 655 CTTTC-----CCA-----TCAGACAGCC-----GAG-----CT-----CGCTCCGGAA 688
Db 1621 AMTSCYKSGAGLCIASTNTDTACVAATAATDAVYLGAGNTSSANCNMKAGCTNNGTTA 1680
Qy 689 CTTGTGATC-----TCCAACGACAAAGAGAGCTTTTGCCACTGACTC-----GGCG 735
Db 1681 CVAKTCAAGAAGITFNHTNCNSYLNTCTVNSGNSACQWASKADQTOASCLYSEGECEV 1740
Qy 736 TGTCTCT-----GAG-----GAC-----ACCTCCAGAAC--GCA 758
Db 1741 VGTSCVRKTCDAATDAATDADDTECSYQOSCTVARLGACQARACATYKSLQCKFNTS 1800
Qy 759 GG-----TGCTGGCGCCCGCTT-----CTGCTGGGACCCCGGAA 793
Db 1801 GSKCFWNPNTKTCVDLNCNIEATTLYDTHNECVAVDATLACTVRATNGABAQCGMARGA 1860
Qy 794 CTTCT-----CCTGCGCG-----AAGCGGAC-----GGC----- 818
Db 1861 CASYTTIEOCKTNASNGVCVMNTNANLPAPACQDKSCTSAPTSTTHNDYVAYNTATVK 1920
Qy 819 -----AGGATGGGCCCCCACTT-----CG-----CCTGCG----- 844
Db 1921 CTVATPNSGGNPTLGGCQTAACSSVIDKEQOINANGDPGCGWGTQCADKSCATASA 1980
Qy 845 -----CCACTTGACT-----TC-----ACCAATCCCT 867
Db 1981 TADYDDDTKCRAYITNKCTVSDSGGCVETPATCETMTQKCYNKGADPCYWTGTACIT 2040
Qy 868 TCTGGAGA--CTAAACCTGCTCT-----C-----AGGAGCGAAGACTGTGAAC 911
Db 2041 KSCDNAPDATADECNVLAGCTLNNVKCKTKCEDPAFDALCKQAISTCTTNGTNC 2100
Qy 912 TT--GT-----AGCTGAGAGC-----CAGAGC--TAGCTCTGGCC 944
Db 2101 VTRGTCFOALSAGCGVSTSTNQCEWIPAVLNASNVITSPAYCTIKNGSTAPITLTSB-- 2158
Qy 945 ACCAG-----CT-----GGGCGAGCTCACCTGTCTCCACCCACC-----CCCA 985
Db 2159 AACAGVFTNCTTKNGGGCVTKSTCSAVTIDVACTTALNGTVCAWDSAQNKCRDKCQDPS 2218
Qy 986 GTT-----CTA--AGGTCTTT-----TCAGAGC--GTGG-----AGGTGTGA 1019
Db 2219 GTTHAACQARAGTAGAGKCARVQNCBQTSVRAACIEGTNGPCLMTDKYQNTDGTGKA 2278
Qy 1020 A-----GGAGT-----GGC----- 1028
Db 2279 CFPYTSCKSLNWNDSCKWISNKCTNGSNCGITLCSETNDDGCVTVGDGACIQSVP 2338
Qy 1029 -----TGCT-----CTCCAACT-----ATGCCAAGCGGC-----GGC----- 1057
Db 2339 DLNSSDPKCKPYTSADAFYTHSDQIASSKCTTNGTTGCIALGSCSSVTVQAGCVFN 2398
Qy 1058 -AGAGCTGTCTTCTGCTCTCCTTGGAGAAAGGTTCCTTGCCTGATTATGAACCTC-- 1114
Db 2399 DKGLTSGVITST--GICTWDTSSCRDQSCADLTGTTHATCSSQLSTCTSDGTTCLL 2456
Qy 1115 -----TATAATAGAGT-----ATAT-----AGGTTT----- 1135
Db 2457 KGACTSYTTQTACTTAVGSDGACIYELASATNNNTAKRELLTCADIQNGTATNVCSVALS 2516
Qy 1136 -----TGTAC-----CTT--TTTTAC-----AGAAAGGTGACT--TTCT-- 1155
Db 2517 TCVSNGTACIPKANGCTVTSKVACNSGGIDGICVFQSTQSTATGAAAGTGCALMTACTVAN 2576

QY 1166 -GTAACAA-----TCGATGTATATAAATTTTATATAA 1199
DB 2577 NDQTACQARDRCSTWASGTRATAVASKATHTCATNOA 2616

RESULT 7
A42125
trophozoite cysteine-rich surface antigen 170 - Giardia lamblia
N:Alternate names: CRP170; cysteine-rich surface antigen CRP170
C:Species: Giardia lamblia
C:Date: 05-Dec-1998 #sequence revision 05-Dec-1998 #text_change 01-Dec-2000
C:Accession: A42125; B42125; S00530; S48056
R:Adam, R.D.; Yang, Y.M.; Nash, T.E.
Mol. Cell. Biol. 12, 1194-1201, 1992
A:Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 gene
A:Reference number: A42125; MUID:92186850; PMID:1545800
A:Accession: A42125
A:Molecule type: DNA
A:Residues: 1-98 <ADA1>
A:Cross-references: GB:M83937; NID:9159124
A:Experimental source: trophozoite
A:Note: sequence extracted from NCBI backbone (NCBIN:88421, NCBI:P:88427); this ORF is not
A:Note: the authors report but do not show 19 tandem repeats of the sequence of residues
A:Accession: B42125
A:Molecule type: DNA
A:Residues: 1259-1766 <ADA2>
A:Cross-references: GB:M83933; NID:9159122
A:Note: sequence extracted from NCBI backbone (NCBIN:88424, NCBI:P:88431); this ORF is not
R:Adam, R.D.; Aggarwal, A.; Lal, A.A.; de la Cruz, V.F.; McCutchan, T.; Nash, T.E.
J. Exp. Med. 167, 109-118, 1988
A:Title: Antigenic variation of a cysteine-rich protein in Giardia lamblia.
A:Reference number: S00530; MUID:88069405; PMID:3335828
A:Accession: S00530
A:Molecule type: DNA
A:Residues: 1154-1409, 'A', 1411-1420, 'K', 1422-1425, 'R', 1427-1481 <ADA3>
A:Cross-references: EMBL:X06741; NID:93355; PID:929603
R:Yang, Y.; Adam, R.D.
Nucleic Acids Res. 22, 2102-2108, 1994
A:Title: Allele-specific expression of a variant-specific surface protein (VSP) of Giardia
A:Reference number: S48056; MUID:94301794; PMID:8029018
A:Accession: S48056
A:Molecule type: DNA
A:Residues: 1-56 <YAN>
A:Cross-references: EMBL:L25059
A:Experimental source: trophozoites WBA6
A:Note: the source is designated as Giardia intestinalis
C:Comment: This translation was produced by PIR staff from information provided by the
C:Genetics:
A:Gene: VSPA6
C:Keywords: surface antigen; tandem repeat

Query Match 11.7%; Score 884.5; DB 2; Length 1766;
Best Local Similarity 24.2%; Pred. No. 1.5e-33;
Matches 423; Conservative 33; Mismatches 636; Indels 655; Gaps 99;

QY 2 ATCTGGGGTG--CTGCCAGGAA-----AAAGCAATTCGGAATT-----40
DB 87 APTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSDSAKE 146
QY 41 -----AATGGTTTGGTGA-----TTTTTAATCTCT-GCTGGCG-----GAG 78
DB 147 CKKNAPCTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSDV 206
QY 79 AGGCGCGCTCTCCCGGTA-TCAGCGC-----TTCTCATTC-----115
DB 207 SDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGY 266
QY 116 ----TTTGAATCCGGGCTCCGGCTTTCCGC-----GTGACAGCCAG 154
DB 267 TDDSVSDAKECKKCNAPCTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSA--VD 323
QY 155 CCGGAG-----GAAGC--CTGTTGCAATTTAAGCGGGCTGTGAA-----C 193

DB 324 CQGSAGYTTDDSVSDAKECKKNAPCTACAG--TADKC-TKCDANGAAPYLKKNPSPDPT 380
QY 194 GCCCAGGCGCGCGGGCGGGCGCGGCGCCATTTTGAATAAAGAGGCGGTGCTTC 253
DB 381 GTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPC-----TACAGT-----ADCKTKC 427
QY 254 CAGCAGGCTCTATAAGTGACCGCGCGG-----CGAGC-----GTGC 291
DB 428 DANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTAD 487
QY 292 GC-GCGTTGCA-----GCTCA-----CTGTAG-----CGGACTTC 320
DB 488 KCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTAC 547
QY 321 TTTTGGTTTCTTCTCTTTTGG-----GGC-ACCTCTGGA-----354
DB 548 ----AGTADKCT--KCDANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAK 600
QY 355 ----CTCCTCC-----CAGC-ATGAAGCGGTG-----AGCCCGGTGC--CGCG-----393
DB 601 ECKKNAPCTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSD 660
QY 394 ----CTGCTA-CGAGCGCGGTGTCTGC-----CTGTGCGAA-C-GCA 428
DB 661 VSDAKECKKNAPCTA--CAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGS 718
QY 429 GTCTGG-----CCATCGCGCG--GGGC--CGAGGA-----AGGCGCGCGCAG 467
DB 719 GYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVD 778
QY 468 CTGAGG-----AGCGCTGAG--CTTGTCTGAC-----GAC 496
DB 779 CQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTC 838
QY 497 ATGAACCACTG-----CTACTCCGCGCTCGGGA-ACTGGTACCCGGA-----538
DB 839 VSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKN 895
QY 539 ----GTCC-----CGAGAG-----GCACTCAGCTTAGCCAGGTGAAATCTTAC 578
DB 896 PSDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTA--ACAG-----TADCKTKC 947
QY 579 AGCGC-----GTCACTGACTACATCT-----CGACCTGCAGTA 613
DB 948 DANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTAD 1006
QY 614 GTC--C-----TGGCGGAGCCAGCC-----CCTGGACCCC 641
DB 1007 DKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTA 1066
QY 642 CTGATGGCCCCCA-----CCTTCCCATCCAGACAGCCGAG-----676
DB 1067 CAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDCQGS-----AGYTTDDSVSDAK 1120
QY 677 ----CTGCTCCGGA--CT-----TGTCACTCTCCACGACAAAAG 711
DB 1121 ECKKNAPCTACAGTADKCTKCDANGAAPYLKKNPSPDPTGTCVSAVDC-----QG 1171
QY 712 GAGCTTTTG-----CACTGACTCGGCGGTGCTCTGACACCTCCAGAC-----755
DB 1172 SAGYTTDDSVSDAKECKKNAPCTA--ACAGTADKCT-----KCDANGAAPYLKKNPSPD 1223
QY 756 ---CGAGGTGCTGGCG-----C--CGTTCTGCTGGGACCCCGGGAAC-----794
DB 1224 PTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLK 1283
QY 795 CTCTCTCTCCCGGAAGCCGGA-----CGGAGG---GATGGGCCCCAATTTCG 838
DB 1284 TNPSDPTGTCVSAVDCQGSAGYTTDDSVSDAKECKKNAPCTAGTATGCTCSCSANCER 1343
QY 839 CCCTGCCCACTTGACTTCAACCAATCCCTCTCTGAGACTAACTGCTGCTCAGGAGC- 897

Db 1344 CDQNDVCARCSITGAPPENGKCPAATPGCHSSCDG-----CTENAMTNQADKCT-----GCK 1394
 QY 898 -----GAAG--GACTGTGAACCTGTA-----GCTGAAGA-----GC--CA 929
 Db 1395 EGRYLKPESIAAGOSGICL-PAEECTSDTHFTYKEKAGDSKGMCLSCSDATHGITGKKCA 1453
 QY 930 -----GAGCTAGCT-CT-----GGCCA--CCAGCTGGGCGAAGTCACTCCCTGTCTCCA 973
 Db 1454 LKTLGSAESTVVCSECTDKRLTTPSGNACLNDNCPAGTYADNINGVSVCAASCHATCAECNG 1513
 QY 974 CCCACCCCAAGTCTTAAGTCTTTT--CAG--GCGTGGAGGTGTGAAGGAGTGC 1028
 Db 1514 DANAASCTACYPGSLYLSGAGTCTVKECTGAFINCAAGQC-----TANVGAKYC 1566
 QY 1029 TGC-----TCTCAAACTATGCAAGGCGCGGCGAG-----AGCTG-----GT 1066
 Db 1567 AQCQKGVAPIDGICTTVAAGRDASVCTAAD-GKCTKCAEYTLMSGGCVGVAKLPGKSV 1625
 QY 1067 CTTCTGTCTCTCTTGGAG-----AAAGTTCTGTTC--CCTGATTATGAACCTTA 1116
 Db 1626 CTLASNGKCTMCAANNOAPVEEKPECESECAKNDNSNACTECLPGYYKAGDKCFKCTA 1685
 QY 1117 TA-----ATAGAGTATAGGTTTGTACCTTTTACAGGAAGGTGACTTCTGTAA 1169
 Db 1686 SSGNNQITGVANCVTCAPAGG--NGPVTCTYKIDGNTGGSVNSKSLSTGAIAGISV 1743
 QY 1170 CAATGCG 1176
 Db 1744 AVVVVVG 1750

RESULT 8

MEGF6 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
 C:Accession: T13954
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030
 A:Accession: T13954
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1574 <NAK>
 A:Cross-references: EMBL:AB011532; NID:G3449293; PIDN:BA032462.1; PID:G3449294
 A:Experimental source: strain Sprague-Dawley; brain
 C:Genetics:
 A:Gene: MEGF6

Query Match 11.4%; Score 860.5; DB 2; Length 1574;
 Best Local Similarity 25.7%; Pred. No. 1.7e-32;
 Matches 364; Conservative 30; Mismatches 552; Indels 473; Gaps 79;

QY 4 CTGGGGTGC-----TGCAGGAAA--AGCAATCTGGAAGTAATGTTT 49
 Db 251 CAEGNG-GOMHICBELRLAHGCHPGYQLAARDKTCEDVDECALGLA----- 297
 QY 50 GAGTCAATTTTAAATCTTGTGGCGG-----AGAGCCCGC----- 86
 Db 298 QCAHCLNTQSFVKVCHAGYELGADGQCRIEIVNSCEAGNGGCHGSHSTSTGPL 357
 QY 87 CTC-----TC-----CC--CGGT-----ATC-----AGCGTTCC 109
 Db 358 CTCPRGYELDEQKTCIDIDCANSFCCQACANTPGGYECSPAGYRLNTDGGCEDVD 417
 QY 110 TCAT-----TCTTTGATCCG-----GGCT----- 130
 Db 418 ECASGHGGEHCHNSIAGSFQCFEAGYRLDEDRGCTSLSESVVLDLGRPLFVRLPHI 477
 QY 131 -----TCTTTGATCCG-----GGCT-----CC-----GGGTCTTCGGCGTC 147
 Db 478 AVLRLDELPRFQDDYCAEEAAAEALRGHTLITEKFVCLDHSFGHDCSLTCDRCRNGTC 537

QY 148 -----AGACCAGCCGAGGAGAGC-----CTGTTTG-----CAATTAAAG 181
 Db 538 FPGQDGDCEPBGWTGIIICNETCPDTPFKNCS9PCTCQNGSTCDPVLGACRCPFVSGAH 597
 QY 182 CGGCT-----GTGAACCCAGGAGGCG-----GGCGGGGCGGGG-----CGAGGCGGGC----- 227
 Db 598 CEDGPGFYGKHKRKKCHANRGRCHRLYGACLDPLGYRPFCHLACPFWAFPGGSEED 657
 QY 228 --CATTTTGAATAAAGAGGCGTG-----CCTTCAG-----GCAGGCTC-----TATAAGTGA 273
 Db 658 CLCEQSHTRSCNPKDGSCKAGPQGERCAECESGFFGFCRHRCTCQCPGACDPVSGE 717
 QY 274 CCGCGCGGCGAGCGTGGCGGCTGAGGTCACTGTAGCGGACTTCTTTTGGTTTCTT 333
 Db 718 CRTCPPPYQGEDCGQCEPVGTFVNCVS--GSCSVG--APCHRVTEGLCLCPKGTGEQGA 775
 QY 334 TCTCTTTGGGCG--ACCTCTGAGCTAC-----TCCCCAGCATGAAGCGCTGAGCCCG 386
 Db 776 DCP--EGRWGLSCQEIACPCEHGA--SCNPETGTCLCLPGFV-----GSRQDT--CSAGW 825
 QY 387 TCGCGGCGTCTACGAGCGGTGTCTGCTGTCTGCGAAGCAGCTCTG-----CCATCGC 441
 Db 826 YGTGCG--IRACACANDGHCDFT-----TGRCSCAPGTGLSCORACDSGHMGPDCIHPNC 879
 QY 442 CGGCGCGGAGGAGGCGCGGCGAGCTGAGGAGC--GCTGAGCTTGTCTGAGCAGCATG 499
 Db 880 SAGHNC-----DAVSLCLCE--AGYEGPRCEQSCRQYVYGPSCQK-----RCEHG 926
 QY 500 AACCACTCTACTCCG-----CCTG-----CGGGAACCTGTTACCCGAGTCCCGA 545
 Db 927 AACDHVSG--ACTCPAGWRGSCFCHACFAGPFGLDCLDSACNCSAG--APCDAVTGSCICPA 983
 QY 546 GAGG--CA-----CTC--AGC--TTAGCCAGGTGGAATCCATCCTACA-----G 580
 Db 984 GRWPRCAQSCFPLTFLGNCQICITCFNGASCDVSTGQCHCAPGMWGTCLQACPPLGLY 1043
 QY 581 CGCGTCATC--GACTACATCTCGAAGCTGAGGTAGTGC-----CTGGCGGAGCCAGCCCT 634
 Db 1044 KNCQSHCLCRNGRCDPILGQCTCPGWTGLA-----CENECLPGHYAAGCQLNCSLH 1097
 QY 635 GGACCCCTGATGGCCCCACCTTCCCATCCAGACAGCCGAGCTCGTCCCGAACTTGTG 694
 Db 1098 GG-----ICDRLTGHCLCPAGWTGDKQSSCVSGT-----FGVHCEHCACRKGASCHVT- 1148
 QY 695 ATCTCCAAAGCAAAAGAGGCTTTTGGC-----ACTGACTCGGCGGTGTC-----TGACA 745
 Db 1149 -----GACFCPPGWRGPHCEQACPRGWFGEACAQRC-----LCPTNASCHVHTGSCR 1195
 QY 746 CCTCCAGAAC--GCAGGTCTGGCGCCCGTTCTGCTGGGACCCCGGGAACCTCTCTCTGC 803
 Db 1196 CPFGFTGLSCBQACQPGT-----FGKDCHEHLCQCFGETWACDPASGYCTC----- 1240
 QY 804 CGGAAGCCGAGCGGAGGAGTGGGCG--CCCAACTTGGCCCTGCCCCACTTGACTTCAACAA 861
 Db 1241 ---NAGYHGTGCLQRCPSGRYPGCEHIC-----KCLNGGTCDPATGAC----- 1281
 QY 862 ATCCCTTCTGTGAGACTAAAACCTGTG-----CTCAGGAGGAGGAGT-----GTGAAC 911
 Db 1282 -----YCPAGFLGADCLACFPQGRFPCSCHVCAACRQSAACDPVSGACICSPGKTGVR 1335
 QY 912 TTTAGCTCTGAAGACCCAGAGCTAG--CTCTGGCCACCAG-----CTGGGCGACGT 960
 Db 1336 ---EHGQDFRFGKGCLEKLCACRNGLCHATNGSCSCLPLGMMWPHCEHACPAGRYGAACL 1392
 QY 961 CACCC-----TG-CTC-----CCACCC-----CACCCCAAGTCTTAAGT 995
 Db 1393 LECFQNNNSCBFTTGAACLCGPFYQACBCEHSCPSGFHGPQRCVCEQQQAPCFVSGQ 1452
 QY 996 CTTTTTCAGAGCGTGGAGGTGTGGAAGGAGTGGC-----TGCTCTCCAAAC-----TA 1042
 Db 1453 CL-----CPAGFHGQCEKGSFSGDCLQCNCHTGVPCDPIISGLCLCFPGRTG 1504

QY 1043 TGC---CAAGCGCGGGCA-GAGCTGGTCTTCTGCTC 1077
Db 1505 AACDLDCRRRFRFG--PGCALRDCGGGADCDPIISGQCHC 1541

RESULT 9
E70806
Hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70806
R: Cole, S. T.; Broach, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S. T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Connor, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: E70806
A: Status: Preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1381 <Col>
A: Cross-references: GB: A102022; GB: A123456; NID: 93261554; PIDN: CAA17744.1; PID: g292444
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv3507
C: Superfamily: collagen alpha 1(IV) chain

Query Match 11.4%; Score 860; DB 2; Length 1381;
Best Local Similarity 30.5%; Pred. No. 1.7e-32;
Matches 418; Conservative 23; Mismatches 614; Indels 314; Gaps 54;

QY 1 GATCTGGGTGCTGCCGAAAGCAAAATCTGGAAGTTAATGCTTTTTCAGTGAATTTT 60
Db 147 GAAGSGGNGSAGLWNGGAGGAG-----GSGGAAGGNGGWLFGAGGTGGTGG 200

QY 61 AAATCTCTGCTGGCGGAGA---GGCCCGCTCTCCCGATATACGCGTCTCTCATCTTT 117
Db 201 APG---AMGCTGGGNGGALLTGG-----GGGAGGAGGCGTGGTGGG 244

QY 118 TGAATCCGCGCTCCGCGTCTTCGCGTCA-GACCAGCGGAGGAGCCCTGTTTGA-- 174
Db 245 GNGALLIAGGVG--GAGG---TGGGTGAGGAGAGGTGGNGAGGLFNGGDDGAGG 298

QY 175 -----ATTAAAGCGG-----GCTGTGAAGCCAGCGCGCGCGCGCGCGCGCC 218
Db 299 QGGDGAAGAAASAGGTGCGGCGGCGTGGAGGAGPVLFGHGGAGGCGGCGTGMG-- 356

QY 219 GAGCGCGCCATT--TGAATAAGAGCGGTCCCTCCAGGAGGCTCTATAAGTGACCG 276
Db 357 GAGGDTTVIAAGTGGGTGAAGAGG--AAGARALTGGGLAGG---VQAGTG--G 408

QY 277 CCGCGCGAGCGTGGCGCGC-GTTGAGGTCACTGTAGCGGACTTCTTTTGGTTTCTTTC 335
Db 409 TGGTGGGADAAVVGFGANGDPGAGGK--GGNGGIGGA-----AVTGG----- 451

QY 336 TCTTTGGGCACCTCTGACTCACTCCCGCATGAAGCGCTGAGCCCGTGGCGCGCT 395
Db 452 ---VAGDGTGGKGTGGA--GGAGNDAGSTGNPFGKGGDGGTGGAGGAGGAAGTGGGHA 507

QY 396 GCTACGA-GGCGGTGTGCTGCTG-----TCGGA-----ACGCACTCTGCGCATC 439
Db 508 GNTGDDGDTGGGNGGTGGVNGADNTLNPDPGAGBPGGAGGAGGAGCAAGPGGTG 567

QY 440 GCCCGGGCGGAGGAAGGG-----CCCGGAGCTGAGAGCGCGCTGAGC 484
Db 568 G--TGGNGGNGGNGGNGGNGGNGAGNNSTNAPVGGEGAGGDDGGAGGAGANGG 625

QY 485 T-----TGCTGACGACATGAACCACTGCT-----ACTCCCGCTGCGGAACCTGG 530
Db 626 TAGSQTGGVGGDGGAGGNGGCGKAGTNSGNFVYDGAAGGAGGNGVGGAGANGG 685

QY 531 TACC CGGAGTCCCG--AGAGGCACCTCAGCTTAGCCA-----GGTGGAAATCTCAGACGC 583

Db 686 TCGSGNGDGGAGGIGGAGGNGIPGTGTGTPAGGTGAKGGDGGGAGGAGNAGAGGQ 745
QY 584 GTCATCGACTACATTCT-----CGACCTGCAGGTAGTCTCTGGCCGAGCAGCCCTTGAC 638
Db 746 CGNAGCGGAGGAGGAGNAVIPGDGVGKAPHGADAGSGDGGKGGGSGGTGSGGAPTGG- 804

QY 639 CCCCTGATGCGCCCGCCACCTTCCATCCAGACGCGAGCTCGTCCGGAACCTTGTCTCT 698
Db 805 ---AGTGG-----SGGHAGKGGAGI-----GAQGTITIVPNG 836

QY 699 CCAACGACA--AAAGAGAGTTTGGCACTGACTCGGCGG-TGTCTGACACCTCCAGAAC 755
Db 837 GNAGDGGNGNAGAGNGSGDFGNTTSGASGSGNGNAGTAGSG-----GA 885

QY 756 CGAGTGTCTGGCGCCGCTCTCCCTGGGACCCCG-----GGAACCTCTCTCTG 802
Db 886 GTGTGTGSGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 945

QY 803 CCGGAAGCCGACGCGCAG-----GGATGGGCCCC-----AATTCGC 839
Db 946 GAPGPAGAPGPTTGNAGSGGIGGGDGGKGGDADAVNVVFMPTPEQAATGTAG- 1004

QY 840 CTGCCCCACTTCACTTCACTCA-----ATCCCTTCTTGAGAG 877
Db 1005 -SAGDPTGGNGGPTGPGSPVAPPPTPTVQCGDGGAGGTGNTANDGTATGKGG- 1062

QY 878 TAAACCTGTCTCAGAGCGAAGG--ACTGTCAACTTCT-----AGCCTGAAG 924
Db 1063 -----EGVSGILGPGGNGGTGNAATGNGVNAAGNGKGGDGGGFGAGGAGG 1116

QY 925 AGCAGAGTAGTCTTGGCC--ACAGCTGGGCGAGGTACCCCTGCTCCACCCACCC 982
Db 1117 SVTDSAGTAGNGGNGNATGTTAGOPAGNGAGGK--GDGNG----- 1161

QY 983 CAAGTTCTAA-GTCTCTTTTACAGAGCTGAGGTGTGAAGGAGTGTCTCTCCTCAAACT 1041
Db 1162 IAAATGTAGNGGNGNGNDGAVNAGTGGSGNG--GNAGGGGANGDGG----- 1209

QY 1042 ATGCCAAGCGCGCGAGAGTGTCTTCTGCTCTCTCTGAGAGAGGTTCTTGTGCCCT 1101
Db 1210 -----ACGAGGAGGCGKIDGGPGDGGNGGNGGTGAGNGGNGGTGGVSGAAG 1263

QY 1102 GATTATGAACCTATAAGAGTATATAGGTTTGTACTTTTACAGGAGGTGACT 1161
Db 1264 DGGNGTGGFAGGTGAGGSGG-GTGGAGGDDGTGG-----DGGNGVIAGGGTGGNGAS 1319

QY 1162 TT--CTGTAACA-----ATCGGATGTAT--ATTTAACTTTTATAAAAGT 1202
Db 1320 GAGGAGGTGCFAGNGAGNGCGTGGASDGDNGNAGSGATGCTGGNGGT 1368

RESULT 10
S09118
G surface protein 168 - Paramesidium primaurelia
C:Species: Paramesidium primaurelia
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 07-Dec-1999
C:Accession: S09118
R: Prat, A.
J. Mol. Biol. 211, 521-535, 1990
A: Title: Conserved sequences flank variable tandem repeats in two alleles of the G sur
A: Reference number: S09118; MUID: 90172419; PMID: 2308165
A: Accession: S09118
A: Molecule type: DNA
A: Residues: 1-2704 <PRA>
A: Cross-references: EMBL: X52133; NID: g10049; PIDN: CAA36378.1; PID: g578473
C: Genetics:
A: Genetic code: SGCS
C: Superfamily: G surface protein

Query Match 11.4%; Score 857; DB 2; Length 2704;
Best Local Similarity 23.8%; Pred. No. 3.3e-32;
Matches 440; Conservative 59; Mismatches 630; Indels 716; Gaps 98;

A;Gene: RV0578c
C;Superfamily: collagen alpha 1(IV) chain

Query Match 11.3%; Score 852; DB 2; Length 1306;
Best Local Similarity 30.8%; Pred. No. 3.7e-32;
Matches 404; Conservative 24; Mismatches 507; Indels 378; Gaps 60;

QY 1 GATCTGGGTCGTCAGAGAAAGCAATTCGGAAGTAATGTTTGGTGGTGAATTTT 60
DB 132 GGILYNGNGGSG--APGCGGAGGAAGLIGNGNGG-----AGGVGTTG---GAGGHG 182
QY 61 AATCTCTGTCGGGAGAGCCGCTCTCCCGCTATCAGCGCTTCCTCATTTCTTGA 120
DB 183 AGWLYNGAGGAFGAGAVGNGG-----AGGTAGLFGVG-----GA 220
QY 121 ATCCGCGCTCCGCGT-CTTCGCGCTCAGACAGCCGCGAGAGAGCTGTTTGAATTTA 179
DB 221 G---GAGGNGIAGVGTSTPSSGCTAGGAGGIGNGGAGGAGVLMG-----NGG 269
QY 180 AGCGGCTGTGAACCCCGG-----CCCGCGCGGCGCGGCGGCGGCGCAT 231
DB 270 NGAAGGGGPGGAGAGAGAHATNLGADQAGGNGNGAGAGTG--GVGGPGG--HGLL 326
QY 232 TTGAATAAAGAGCGCTCCCTCCAGGCA--GGCTCTAT-----AAGTACACCCCGCGC 283
DB 327 GLGSHGAGGAGGSGD-----GGAPGCGNGATGTWGNLGAAGTGNGGNPAGGA 379
QY 284 GAGCGTGGCGC-CTTCAGCTCACTAGCGGACTTCTTTGGTTCCTTTCTTTG 342
DB 380 GGAGGASVGSAGANGAPGTTSTGGNGGDKGAD-ALSSQGTGANGRGDGGQVGN 438
QY 343 GGCACCTCTGACTCACTCCCGAC-----ATGAAGCGCTGAGCCCGGTGCGCGCTGC 397
DB 439 GGAGGAGRGGA-----GGLGFGSEAPRPGAGGTGGAGNGGTQAGDGTGG- 487
QY 398 TACAGGCGGTG-----TGCTGCTGTGCGAACGAGCTCTGCGCATCGCCCGG 446
DB 488 AGAGGDDGSGGAGSIFNAPAGAPGNGGNGGPGGAGG-EGGAGGLAALASQN 545
QY 447 GCCGAGG-----CAAG-----GGCCGCGGAGCTGAGGAGCGCGTGA 482
DB 546 GSQAGGDDGAGGNGTTPGNGCHGAALGVNGVGG--AGHGGDFGVGAG--GQGS 601
QY 483 GCTTGCTG-----GACGACATGAACCACTGCTACTCCTCCCGCTCGGGNACT 528
DB 602 GSTPGANGAPNTPTSGNGNGRGADATG---FGQTGASGGRGSDGLVNGGAGGAG 658
QY 529 G-----GTACCGGAGTCCGAGAGGCACTCAGCTTAGCCAGCTGGAATCCT 576
DB 659 GNGSKLPGLRLGNPLDGTGG---NGHAGSGGHWAGNGTG--GAGGTGGVGTGS 714
QY 577 ACAGC-----CGTCACTGACTACATTCGACCTGACGAGTAGTCT 618
DB 715 GSDGVNGSSAGADGHPGTGGVGTGGKGGDGGGAAPNGVAGSQGPGGAGGDDGTGGV 774
QY 619 GCGGAGCGAGCCCTGACCCCTGATGGCCGCCCATCTCCCATCCAGACAGCCGAGCT 678
DB 775 GGNNGRG-ITGADGATAGA-----RGQGG-----AGHGGKGRGT 811
QY 679 CGCTCCGGA--ACTTGTGCTATCTCCACGACAAAGAGACTTTGCCACTGACTCGGCCGT 736
DB 812 GG---PGGAPAGTTG-----SQAGGNGSGGTGGTGGDGGNGANGSVFT 854
QY 737 GTCCTGACACCTCCAGAGCGAG--GTGCTGGCGCCCTCT-----GCTCGGA 784
DB 855 NNGIGG-----NGNGNGNAGPSGAGSGGAGGTGATGSSSIHNGNGNGNGND 906
QY 785 CCGCGGAACTCTCTGCGGAGCCCGG---ACGGCAGGATGGGCGCCCACTTCGCC 840
DB 907 HALSNGAA-----GNGGNGNGSLRGSGGAGGNGGN----- 942
QY 841 CTGCGCACTTGACTTCAACAAATCCCTCTCTGGAGACTAAACCTGGTCTCAGGAGCGAA 900

943 -----ASRG-----MGDDG-----TGAGGN-AGQINGGA 968
QY 901 GGACTGTGAACCTTACCTGAAGAGCCAGAGCTAGCTTGGCCACAGCTG-GGCCAGC 959
DB 969 GGN-GSDGTTGSDNPGAITGSGGRGDDGVGGGGSVAGDAGDGGAGGAGTGTGLRG 1027
QY 960 TCACCTGCTCCCAACCCCAAGCTTCTTAAGTCTCTTTTACAGAGCGTGGAGGTGGA 1019
DB 1028 T-----TGAT-----CATGTFDAGDG-----HGNGGTGGVG--GTGGA 1060
QY 1020 AGGAGTGGCTCTCTCCAAACTATGCCAGCGCGCGGAGAGCTGGTCTTCTGCTCTCT 1079
DB 1061 GGGGNGGAGG-----KALSPGNNGSQAGGDDGAGGAGGTGG-----TGG----- 1102
QY 1080 TGGAGA-----AAGGTTCTGTGCCCTGATTTATGAACCTCTATAATAGAGTA- 1126
DB 1103 DGGRAHGTLPSSLAGTGGTGGTGGT--GTGGAGGAGTGTGATGATCAAGRAG 1160
QY 1127 TATAGTGTGTTTACCTTTTACAGGAAGTGAATTTCTGTAAACAAATGCGATG 1179
DB 1161 NGVGGSGGLGSAFGPGGT-----GGMGGAG-----GTSTVSAGDGGRG 1200

RESULT 12
A;0869
hypothetical glycine-rich protein RV2490c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70869
R;Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genom
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70869
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1660 <COL>
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16067.1; PID:g2791
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: RV2490c
C;Superfamily: collagen alpha 1(IV) chain

Query Match 10.6%; Score 800.5; DB 2; Length 1660;
Best Local Similarity 29.5%; Pred. No. 9e-30;
Matches 393; Conservative 18; Mismatches 551; Indels 369; Gaps 56;

QY 1 GATCTGGGTCGTGCC-----AGGAAAAAGCAAAATCTG-GAAGTTAATGTTTGGTGTG 54
DB 134 GGLLFNGNGNAGAVGVGGAGGAGLFGTGGAGGAGAGAPGTGGTGGWLAGGGVG 193
QY 55 ATTTTAAATCTTGTGTCGCGAGAGGCCCTCTCCCGGTATCAGCGCTTCTCAT 114
DB 194 G-----MGAGG-GAGG-----AGGNAGLFGNG----- 215
QY 115 CTTTGAATCCGCGCTCCCGCTCTTC-----GGGTCCAGACCAGCCG---GAGGAAGCC 166
DB 216 ----GAGGAGGAG-----GAGGAGNAGWFGHGGAGGVGVGAANGATPGDGAAG-V 266
QY 167 TGTTCGAATTTAAGCGGCTGTGAACGCCCAAGGCGC--GGCGGGGCGGGCCGAGGGC 224
DB 267 AGSDDG-AGGDLGAGSDGDDGAGGVGGNGRGMLLNGGAGGAGGVGGVGGAGGAAGA 325
QY 225 GGCATTTTGA--TAAAGAGCGCTGCTTCCAGGAGGCTCTATAAGTGACCGC----- 277
DB 326 GGAGATGINGPAGISAGGDDGAG-----GNGAGNGGVGGAGGAGSAGLLGYVG 377
QY 278 -CGCGCGAGCGTGCCTGCGGCTTGCAGGTCACTGTAGCGGACTTCTTTTGGTTTCTTCT 336
DB 378 RADGGAGGGGLGAGPAGGAGGNGGSLAAGDGGAGG-----HGGDP-----G 422

Db 833 ERADGA-----GPNANGENGSGNGDGGAGNGGAGGAKQAAGYTDGAT 882
 Qy 639 CCCTGATGCCCCCACTTCCATCCAGACACGCA---GCTCGCTCCGGAACCTGTCA 695
 Db 883 GTGGDGGNG---DGGKADGGAGENGSLNSGAMLPFGG--TVGNPG 923
 Qy 696 TCTCCAACGACAAAGGAGCTTTTCCACTGACTCGCCGT-----GTCCTGACACCT 748
 Db 924 T-----GNGGNGGAG--VGTGKAGTGSUT--GLDGTGDTPNNGNGNGNGNGKG 973
 Qy 749 CCAGAACGCGAGGTGCTGGCCCGCTTCTGCTCGGACCCCG---GGAACCTCT 798
 Db 974 GTAGNGSAGAGGNGSGGLNGDAGNG--GNGGALNQAGPFGTGGKNGG---1025
 Qy 799 CTGCGCGAAGCGGACGAGGATCGGCCCCCACTTCCGCTCGCCACTTGACTTCA 858
 Db 1026 -----NGGAGMNGSLGGF--GGAGGG-----AV 1048
 Qy 859 CAAATCCCTTCTCGAGACTAAACCTGTGCTCAGG--AGCGAAGGA-----CTGTGAAC 911
 Db 1049 DVAAT-----TGGAGNGGA---GGFAS TGLGGPAGGAGGAGDFASGVGVGGAG 1097
 Qy 912 TTGTAGCCTGAAGCCAGAGCTAGCTCTGGCCACAG--CTGGCGACCTCACCTGCT 969
 Db 1098 GDGAGGVGGFGGQGGIGGEGTGG---NGSGDGGGGISLGGNGGLGNGGVSETFG 1154
 Qy 970 CCCACCCACCCCAAGT-----CTAAGTCTTTTTCAGAG-----1005
 Db 1155 -----GAGNGGYGPGPGGNGGNGGAGGNGGVSTTGGDGGAGKGG 1200
 Qy 1006 -----CTGGAGGT-----GTGAAGGATGCTGCTCTCCAACTA-- 1042
 Db 1201 NGDGGNVGLGDAGSGGNGGIGTDAGAGGAGGAGGNGGSSKSTTTTGNAGSGAGG 1260
 Qy 1043 ---TGCCAGCGCGCGCAG--AGCTGCTTCTGCTCTCTGAGAGAGGTTCTGTGC 1098
 Db 1261 NGGTGLNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1314
 Qy 1099 CCGATTATGAATCTATATAG-----AGPATATAGTTTGTACCTTTTACAG 1151
 Db 1315 GNGSSGSAAGSG--VNVVTAHGNGGNGGNGGAGGAGGAGGAGGAGGAGGAGG 1373
 Qy 1152 --GAAGTGACTTCTGTAAATGCGATATATATAA-----CTTTTATAAAG 1201
 Db 1374 GDGNGNGGNGSGNSTGVAGLAGAAGAGGNGGTSSAAGHGSGSGSGGTTGGAGAG 1433

RESULT 14
 E70917
 Hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: E70917
 R: Cole, S. T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A: Authors: Scares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: E70917
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-1329 <COL>
 A: Cross-references: GB: Z95844; GB: AL123456; NID: g3250713; PID: CAB09271.1; PID: g2131046
 A: Experimental source: strain H37RV
 C: Genes: Rv1450c
 C: Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 10.5%; Score 789; DB 2; Length 1329;
 Best Local Similarity 29.3%; Pred. No. 2.6e-29;

Matches 378; Conservative 21; Mismatches 523; Indels 368; Gaps 48;
 Qy 7 GGGTCTCCAGGAAAGCAAAATTCGGAAGTTAATGGTTTTCAGTGAATTTTAAATCC 66
 Db 284 GHHGAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGH 327
 Qy 67 TTGTGCGGGA--GAGGCGCGCTCTCCCGGATATCAGCGCTTCTCTCA--TTCTTTGAATCC 124
 Db 328 --GWLAGAGGAGGAGG-----VGGTGGAGGAGFSRLLIVAGDNGDGFAGGAG 373
 Qy 125 CGGCTCC---CGGCTTTCGGCTCAG---ACCAGCGGAGAGACCTGTTTGAAT 176
 Db 374 GTGAGAGSTIGHAGAGASPTSGGNGAGAGNAHFSSGKAGGNGAGG-----AGLGVN 428
 Qy 177 TTAAGCGG--GCTGTGAAAGCGCCAGCGCGCGCGCGCGCG-----GGCGAGCG--GGCC 228
 Db 429 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 488
 Qy 229 ATTTTGAATAAGAGG--GTGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 279
 Db 489 ATGATGTLNGLAGAGADGTGGKGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 548
 Qy 280 CGGCGAGCGTGC-----GCGCGTTGAGGTCTCTGTAGCGGACTT-----CTTTT 324
 Db 549 AGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 608
 Qy 325 GGTTCCTTCTCTTTTGGGCGACCTCTGGAATCTACTCCAGCATGAAGGCGCTGAGCCC 384
 Db 609 GGGGT-----GGGANATVAGG-----AGGAGGAGGAGGAGGAGGAGGAGGAG 646
 Qy 385 GGTGCGCGCGTCT-----ACGAGCGGTG--TGCTGCTCTCGGAAAGC 427
 Db 647 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 706
 Qy 428 AGTCTGGCCATGCGCCCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 484
 Db 707 NATTPGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 762
 Qy 485 TTGCTG--GACGACATGAACCTGCTACTCCGCGCTCGG--GAACCTGATCCCGGAGTC 541
 Db 763 LAGDGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 805
 Qy 542 CCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 601
 Db 806 PGAGAGHGTPTSGG-----DGGDGGGNGGNSQVVGNGGDDGG-----842
 Qy 602 GACCTGCGAGTGTCTCTGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 646
 Db 843 ----NGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 892
 Qy 647 GGGCCCCCATCTCCCATCCAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 706
 Db 893 GG-----AGLNGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 922
 Qy 707 AAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 766
 Db 923 GVAVGAGP--QPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 968
 Qy 767 CCGCGCTTCTGCTGCGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 826
 Db 969 DGAATPDGANG--GNGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1012
 Qy 827 GCGCCAACTTCCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 886
 Db 1013 -----TG--STAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1047
 Qy 887 TGCTCA---GGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 942
 Db 1048 PGHGGNGGNGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1107
 Qy 943 CCACGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1002
 Db 1108 QGLTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1143

QY 1003 GACGCTGG-----AGGTGTGGAAGAGTGGCTGCTCTCCAAACTATGCCAAGCGCGGCA 1058
 Db 1144 GGGGGRGGADAGGDDAGLGGSGPGGTGPDWG-----TGCTGTGTGGTG 1186
 QY 1059 GAGCTGCTCTCTGCTCTCTCCAGAAAGGTTCTCTTCCCTGATTTATGAACCTATA 1118
 Db 1187 GQANGG-----LTGG-----RGTTGNGNGNTGTTGGAGGTG-----GTGHNQSPGNG 1232
 QY 1119 ATGAGG-----TATATAGTGTGTTGTAACCTTTTATACAG----- 1151
 Db 1233 GNGAGGFGNGFAGVGGRGMGSGGTGTGTDAGPFGTGTGTTGGTGGGGGGGFSILL 1292
 QY 1152 --GAAGTGAAGTCTCTGTAACAATGCGATG 1179
 Db 1293 GLGGLGLGSPGSIATGTRAGAGGGGGFGG 1322

RESULT 15
 H70987
 hypothetical glycine-rich protein Rv1759c - Mycobacterium tuberculosis (strain H37RV)
 N;Alternate names: wag22 antigen homolog
 C;Species: Mycobacterium tuberculosis
 C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C;Accession: H70987
 R;Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, K.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A;Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A;Reference number: A70500; MUID:98295987; PMID:9534230
 A;Accession: H70987
 A;Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-914 <COL>
 A;Cross-references: GB:295890; GB:AL123456; NID:G3242245; PIDN:CAB09322.1; PID:G2131027
 A;Experimental source: strain H37RV
 C;Genetics:
 A;Gene: Rv1759c; wag22
 C;Superfamily: elastin

Query Match 10.4%; Score 784.5; DB 2; Length 914;
 Best Local Similarity 29.5%; Pred. No. 3.4e-29;
 Matches 337; Conservative 18; Mismatches 386; Indels 401; Gaps 48;

QY 1 GATCTGGGGTCTGCTCCAG-----GAAAGCAATCTCGAAGTTATGTTTGTAGTGA 55
 Db 116 GNGANGAPGTGANGGDAGLWLTGNGAGGSGAKGA---NGGAGGPGGAAGLFGNGAG-GA 171
 QY 56 TTTTAAATCTCTGCTGGGAGAGGCCGCTCTCCCGGTATCAGCGCTTCTCATTC 115
 Db 172 GGTATA-----NNGIGAGG-----AGSAMLFGAG----- 197
 QY 116 TTTGAATCCGGGTTCGCGTCTTCGCGTCAACAG-----CCGAGAGAGCCT 167
 Db 198 ---GAGGAGGAATSLVGGIGTGTGTGNNAGLAGAGAGGAGGFFSFTAGGAGAGG--- 251
 QY 168 GTTTCATTTAAGCGGCTGTGAACCCAGG--GCCGCGGGGGGGGGCGAGCGG 225
 Db 252 ---AGLFTTGGVGGAGGQHTGGAGGAGGAGGLFG-AGMVGAGGFFDHTLTGTGAGG 307
 QY 226 GCCATTTTGAATAAGAGCGTGCCTTCCAGCGAGGCTCTATAAGTCAACCGCGCGCGCA 285
 Db 308 -----DGGGGLFGAG-----DGGAGGSLTGTGAAGNNGNAGTSLGAAG- 349
 QY 286 GGTGCGCGGCTTGCAGGTACTGTAGCGGACTCTTTTGGTTTCTTCTCTTTGGGGC 345
 Db 350 -----GAGGTGAGGTTFGGGKGAGGA-----GNN- 375
 QY 346 ACCTCTGACTCACTCCCGAGCATGAGGCGCTGAGCCGCGGTGCGCGCTGCT--ACGAG 403
 Db 376 -----AGMLFGSGGGGTGGFAGAGGCG--GVGSSAGMLSGSGSG 415

QY 404 GCGGTGTGCTGCTGTTCGGA--ACGCAGTCTGCTCCATCGCCCGGGGCGGAGGAAA---G 457
 Db 416 GAGSGG-GPAGTAAGAGAGGAGAPELLG--NGNGNGNGESGGTGGVGGAGNAVILNG 472
 QY 458 GCGCCCGGAGCTGAGGAG--CCGCTGAGCTTGC-----TGGACGACAT 498
 Db 473 GEGGIGALAGKSFGFGGGLLLGADGYNAPESTSPHNLQDILSFINEPTALTGRIPLI 532
 QY 499 GAACCACTGTCTACTCCCGCTGCGGAACTGGTACCGGAGTCCGAGAGGCACTCAGCT 558
 Db 533 GNGDSCTPG-----TGDDGA--GGWLFNGZNG--GAGAAGTNGSAGGAG 574
 QY 559 TAGCCAGGTGAAATCCTACAGCGCTCATCGACTACATCTCGACCTGCGAGGTAGTCTCT 618
 Db 575 GAGGILFTGGAG-----GAGVGATAGA-----GGAGGAGSAGFL 609
 QY 619 GCGGAGCGCCCTGAGCCCTGATGCGCCCACTTCCATCCAGACAGCCGAGCT 678
 Db 610 IGSGTGTGVGAATTTGG-----VGAGG-----NAGLLIAGALG-- 645
 QY 679 CGTCCGAACTTGTCTATCTCAACGACAAAGAGAGCT-TTTGCCACTGACTCGCCGTG 737
 Db 646 ---GCGGGAFTAG-----VTTGGAGGTGGAAGLFGANGA---GGAGGTG 683
 QY 738 TCCTGACACCTCCAGAAAG--CAGGTGCTGGCGCCCTTCTGCTGGGACCCCGGAACC 795
 Db 684 STAGAGG-----AGGAGGLYAHGTTGPGNGSGSTAGGTGGAGGPGGLYAGGS--- 734
 QY 796 TCTCTGCGGAGCCGAGCGGAGGATGGGCCCACTTTCGCCCTGCCCACTTGACTT 855
 Db 735 -----GGA-----GGHGGVAGGGGVGG-----NAGSLTL 759
 QY 856 CACCAATCCCTTCTGTGAGACTAAACCTGTGCTCAGGAGCGAGGACTGTGAACCTGT 915
 Db 760 NA-----SGAG-----GSGSLSLGGKAGAGGAG-----GS 785
 QY 916 AGCCTGAAGAGCCAGAGCTAGCTCTGGCCACAGCTGGGCGACGTCACCTGTCTCCACC 975
 Db 786 AGLFVGGGAGGNGYSLNG---TGG-----DGGTGGAGQ----- 817
 QY 976 CACCCCAAGTTCTAAGTCTTTTTCAGAGCGTGGAGGTGTGGAAGGAGTGGCTCTCTC 1035
 Db 818 -----ITGLRSFGF-GAGGAGGASDTGAGNGGAGGKAG----- 850
 QY 1036 CAAACTATGCCAAGCGCGCAGAGCTGTCTTCTGCTCTCTTGGAGAAAGGTCTCTGT 1095
 Db 851 -----LYNGGDGGAGGAGGATSGKGGAGGNVVIENG-----GNNGNAGKAGGTAGAGG 899
 QY 1096 TG 1097
 Db 900 AG 901

Search completed: May 7, 2004, 15:10:51
 Job time : 53 secs


```

QY 841 CACCGACTGAGATATGTTTACATAGTCTG-----TGGGCTGTTTTTTTGTATT 895
Db 1070 LGGPGNGTGNATGNTGNGVANGNGKGGGQGGQGGAGNGGAGGSGVTDGSAGTGN 1129
QY 896 AARCAAAATATT-----TAGATGTGAAAAAA 923
Db 1130 CGNGGNATGTTAGQAPAGNGSAGGKGDDGGNIAAGA 1166

RESULT 15
O88281 PRELIMINARY; PRT: 1574 AA.
ID O88281;
AC O88281;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MEGF6.
GN MEGF6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=983030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RL like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
DR EMBL; AB011532; BAA32462.1; -.
DR PIR; T13954; T13954.
DR HSP; P00716; IAP0.
DR GO; GO:000509; F:calcium ion binding; IEA.
DR GO; GO:0005198; F:structural molecular activity; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00008; EGF; 20.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00179; EGF_CA; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00022; EGF_1; 23.
DR PROSITE; PS01186; EGF_2; 23.
DR PROSITE; PS01187; EGF_CA; 5.
KW EGF-like domain.
SQ SEQUENCE 1574 AA; 165445 MW; 2B48533D8F77F6E7 CRC64;

Query Match 13.3%; Score 764; DB 11; Length 1574;
Best Local Similarity 28.9%; Pred No. 3; 4e-35;
Matches 302; Conservative 31; Mismatches 40; Indels 312; Gaps 74;

QY 1 GGGGCCCATCTCTGTT---TCAGCCAGTC---GCC--AAGATCATGAAGTC--GC--- 46
Db 549 GWTGIIICNETCPDPTFGKNCSSPC--TCQNGTCDPVLGACRCPPGVSGAHCEDCPKGF 606
QY 47 -----CAGTGC---AGCAC-----C-----GCCACGC----- 67
Db 607 YGKHCKKCHCNRRGRCHPLYLACLDPLGLYGRFCHLACPPWAFPGCSSECLCEQSHTR 666
QY 68 -C-----GCCGCGGC-----C-----CCAGTGTGSC--GCTGAAGCGCGGCAAGACAGCAGC 113
Db 667 SCNPXKDGSCSKAGFQGERCQAECSGFFPGCRHRCCTCPGVACDPVSGECRTQCPPGY 726
QY 114 GRTGCGGCGAGGTGTGTGCTGTCTGTCTGAGCAGAGCG--TGCCCATCTCGGCGTGC 171
Db 727 QGEDCCQEC---PVGTFGVNCSSCS-----CVGAPCHRVG-----ECLCPGKT--- 769
QY 172 GGGGCGC-CGGGCGCGC---CTGCTGCC---TG---CTGGACGAGCAG-CAGGTAAA 220
Db 770 -GEDCGADCPGRWGLGCGEICPAGEHGASCNPETGTCLCLPFGVGRCPQDTCAGW--- 825

```

```

QY 221 CGTGC-TGCTCTACGAC-ATGAACGGCTGTACTCAGC-----CTCAGGAG 266
Db 826 YGTGQIRACANDGHCDPTTGRCSAPGWTGLSCQRACDSGHWPDCIHPNCSSAGHN 885
QY 267 TGTGTGCC-----C-AC-----CCTGCCCCAGAAC-----CGCAAGGTGA 300
Db 886 CDAVSGCLCEAGYEGPRCSQYGFSPSCQKCRCEHGAACDHVSGACTCPAGWRGS 945
QY 301 -----GCAAGTGGAGATTCTCCAGCAGTCATCATCATCAGGAGCCTTCTGTTGGAG 356
Db 946 FCEHACPAFGFLDCDSACNCSAGAPCD--AVTGSC-ICPAGRWGPRCAQSCPLTFGLN 1002
QY 357 CTGAATCGGAATCCGAAGTTGGACCCCGGGGGCC-----GA-GGGCTGCC-----GG 405
Db 1003 CS-QICTCFNGASCDSVT-----GQCHCAPGWMGPTCLQACPFGLYGNQHSLCRNGG 1056
QY 406 TC-CGGGCTC-----CGCTC-----AGACCTCTAACCG-----CGAGAT 439
Db 1057 RDPILGCTCPGWTGLACNECLPHYAAGQNLNCSLIH-GGICDRLTGHCLCPAGWT 1115
QY 440 CAGC-GCCCTGACCGCC-----GA-----GGCGCATG-----CGTTCTCG-CGGA 478
Db 1116 GDKQSCSVSGTGFVHCEHCACRKGASCHHVTGAC-FCPPGWRGPHCEQACPRWFGEA 1174
QY 479 CGATGCG---ATCTTGTGTGC-TGAAGCG-----CTCCCCCAGGACCGC 522
Db 1175 CAQRCLCPTNASHHVTGECRCPGFTGLSCEQAQCPGTFGKCEHLCCQCPGETWACDPA 1234
QY 523 GGACCCAGCCATCCAGGGGCAAGAGAAATTAGTGC--TCTGTGGGTC---TCCCCCA 577
Db 1235 SGVCTCAAGYHGTGCL---QRCPG-----RYGPGCEHICKLNGGTCDPATGACYCP 1284
QY 578 ACGGCG-CTCGCCGGAATCTGAGGAGAGAACAGACC--GATCGCGGCCAC---TGCGC- 629
Db 1285 AGTIGADCSLAC-----PGRFSGCAHVACRQGAACDPVSGACICSPKGTGVRCE 1336
QY 630 --CCTTAATCATCCAGCCTGGGCTGA-GGCT--GAGGCACTGGCGAGGAGAGCGG 683
Db 1337 HGPCQDRFGKGCLEKCA--CRNGGLCHATNGSCSCLPLGWMGPHCEHACPAGRYGAACLL 1394
QY 684 CTC-----CTCTCTGCAC-----ACCTACTAG-----TCACCAG-----GACT 717
Db 1395 CFCQNGSCPTTACLGPGFYQAACEHSCPSGPHGPGQRCVCECQOQAPCDPVSGQL 1454
QY 718 TTAGGGGTGGGATCCCATCGTGTGTTTCTATTTTGAAGAGCAGACATTTTAAAAA 777
Db 1455 CPAGFPHG-----QFCEKGCESGFGDGLQQCNCHTGVPCDPIISGLC-----LCPPGR 1502
QY 778 TGTTCACGTTTG--GTGCT-TCTCAG 800
Db 1503 TGAACDLDCRRGRFPGCALRCDGG 1528

```

Search completed: May 7, 2004, 15:19:15
Job time : 57 secs


```

QY 110 -GAGCGGTGGG-----GCAGGTGTGGCTGTCTGTGACGACGAGCGTGGCCAT 160
Db 222 GNGGDDGGAAGAVGISGATGAGDGGHGTGAAGNGGTGGAGSGIDGVGGGTG----- 277
QY 161 CTGCGCTGCCG-----GCGCGCGGGCGCGCTGCTGCTGACACGAGCAGCAG 215
Db 278 -TGNGNGNAGAGGADGSGNSGNG-GIG-----GKGNAGAGGAAGNG 323
QY 216 GTAAAGTGTCTCTACGACATGAACGGCTGTACTACGCTCAAGAGCTGTGCCCC 275
Db 324 GTVGANGTGGDGNNGGAAGAAATAGSNGAGTG-----SAGNGGTGGG----- 367
QY 276 ACCTGCCCCAGAACCGCAGGTGAGCAAGTGA----- 310
Db 368 ---GSGGAGGDTGSGVGGHKGNGADGEVCGAGCGSGPNTSPGNGGCGGSGGAG 424
QY 311 GATTCTCCAGCAGCTCATCACTACATCAAGGACCTT-----CAGTTGAGCTGAATCTCG 365
Db 425 GAACAGAGGAGGAGTAGCGGAGGTGCGAGAAASATNGSGGAGGTGGDGGSGGAGGTG 484
QY 366 GNAATCCGAGTGGACCCCCCGGGCGGAGG-GCTGCGGTCCGGCTCCGCTCAGCAC 424
Db 485 GAGGTGAAGDGGGGGGAGGAGGAGGAGGAGGTGGNGNTGGTAGTAGAAGNGGA- 543
QY 425 CCTCAACGGCGA-GATCAGCGCCCTGACGCCCGAGCGGCATGCGTTCTCTGCGGACATC 483
Db 544 ---AGKGGAGGCGTGGTG--GCGGAGGCGAGGTGGDRTVGGGTVPAGSGGCGGNAG 597
QY 484 GCAT-----CTGTGTCG-----CTCAAGCGCTCCCCCAGGAGCC 519
Db 598 GGGAGGGGADGSGGGDAGTGGNGNGNNGNNGTGGAGN-----GGGGAN 648
QY 520 GCGGACCCCAAGCCATCAGGGGCAAGAGGAAATTA--CGTGTCTGTGGGTCTCCCCCA 577
Db 649 GGAG-----AGSGGGTGGNGAGGAGDAGDAGNGNGTNG----- 686
QY 578 ACGCGCTCGCGGATCTGAGGAGACACAGCCGATCGCGGCCCATCGCGCCCTTAAC 637
Db 687 -----GNGSNGIAGMNGNGAGGTGSG--NGNGSGGNGGAGMNGNSGTGSGD 734
QY 638 GCATCCAGCC---TGG-GGCTGAGGCTGAGGCACTGGCG----- 672
Db 735 GGAGNGGAAGTGTGGDGLTGTGGTGGSGGTGGDNGNGGADNTANMTAAGAGDGN 794
QY 673 -----AGGAGAGGCGCTCTCTGCAACCTACTAGTACCAGAGACTTTAGGGGT 726
Db 795 GGDGFGGAGAGG-GLTAGANGTGGG-----GAGDGGNGATGGHGPJTDPPGNGGT 849
QY 727 GSGATT-----CACTGCTGTCTTCTATTTTGT-----AAAAAGCAGACATTT 770
Db 850 GNGGTGTGTGAGTGLGGLGGTGGDNGNGGTGEGGEGVCGAGGTGGAAGNGDGGTGG 909
QY 771 TAAAAAATGGTCACTT-----TGTTGCTTCTCAGATTCTTCAGAGAAATGCTTT 820
Db 910 TGGDGGAGGTGCTGTTGGLGDPVVGSGDGGT-----GSGGAAGNGNGNAGAGN 964
QY 821 GTATTGTATTACAACTACACCGACTAGAA--TATTGTTTACAAATAGTTCTGTGG- 877
Db 965 GNGGTGAGGIGTGGNGD-ABPVPFAGAGAGAGTTGKGTGGNGSGTGGSGTGGD 1023
QY 878 ---GGCTGTTTTTTTATTAAACAAATAATTAGATGCTGAAAAA 921
Db 1024 GGTGGGNGGTGMWNGKGTGSDTGGGGAGDGGKAPAGGTGGAGDGA 1070

```

RESULT 13

```

Q8VIZ1 PRELIMINARY; PRT; 1384 AA.
ID Q8VIZ1
AC Q8VIZ1;
DT 01-MAR-2002 (trEMBLrel. 20, Created)
DT 01-MAR-2002 (trEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (trEMBLrel. 24, Last annotation update)
DE PE_PGRS family protein.

```

GN MT3612.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxID=1773;
[1]

RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Deicher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.",
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007163; AAK47970.1; --
DR TIGR; MT3612;
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000084; PE_region.
DR InterPro; IPR002173; PFKB.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSELL.
DR PROSITE; PS00583; PFKB_KINASES 1; 2.
SQ SEQUENCE 1384 AA; 110839 MW; 124AC8AAADD55EBD CRC64;

Query Match 13.3%; Score 765.5; DB 16; Length 1384;
Best Local Similarity 31.4%; Pred. No. 2.5e-35;
Matches 349; Conservative 20; Mismatches 469; Indels 273; Gaps 44;

```

QY 1 GGGGCCCATTTCTGTTTTCAGCCAGTCGCCAAGATCATGAAGTCG-----CCAGTGGCA 54
Db 141 GCGN---APGAAGSGGNGSAGLWNGAGAGAGSGGAGNGNGGWLFGAGTGGIG 197
QY 55 GCACCGCCACCCCGCCCGC-----GGGCCCGCAGCTGCGCGCTGAAGCGCGCA---AGAC 106
Db 198 GTGAPGAMGTTGNGNGALLIGGGLGAGGCGGTG--GGTGGTGGNGGALLIGAGGV 256
QY 107 AGCGA-GCGGTGCGCGCGAGGTGGTGCCTGT-CTGTCTGAGCAGAGCGTGGCATCTCG 164
Db 257 GGAGIGGGTGAAGAAAGGTGGNG--GAGLFWNGDGGAGGGG-GDGAAGDAASAGG 314
QY 165 CCGTCCCGG-GCGCGCCGGG---CGCGCTGCTGCTGCTGTCGACGACGACGAGTAAA 220
Db 315 TSGKGGCGDGTGGAGGAGPVLFGHGGAG-----MGCGGTGNGGAGGGTTVIA 367
QY 221 CGTGTCTCTACGACATGAACGGCTGTTACTCAGCGCTCAAGGAGCTGTGCCACCC- 279
Db 368 AGTGGGGTGGAGA---GGAAGARGALTGGLAGVGAGGTGGTGGNGADAATAAVG 424
QY 280 -----TGCCCCAGAACCGCAAGGTGACAAAGGTGGAGATTCTCGACGACGTC 326
Db 425 FGAGNDPPGAPGKGGNGGTGGAAVTGGVAGDGGT-GKGGTGGAG-----AGNDAGST 477
QY 327 AT-----CGACTACATCAGGACCTTCAGTTGGAGCTGAATCG----- 365
Db 478 GNPFGKGGDGTGGAGGAGGAAGTGNHGAGTGGDGGTGGNGGTGGVNGADNTLN 537
QY 366 ----GAATCCGAAGTTGGACCCCGGGGGCCGAGG-GCTGCCGTTCGGGGTCCGCTCA 420
Db 538 PDTGAGGEPGAGGAGGAA-----GPGGTGTTGGNGGNGGNGGNGGNGNG----- 588
QY 421 GCACCTCAACGGCGGAGATCAGCGCCCTGACGCGCGAGCGCGCATG-----CGTTCTGC 475
Db 589 -----GNAGNNSINA PVG--GEGGAGDGGAGGAGGAANGGTAGSGGTGGVGD 635
QY 476 GGACGATCGCATTTGTGTGCTGAAGCGCTTCCCCCAGGACCGCGGACCCCGGACCCAT 535
Db 636 GGAGGNGGGKAGTGNNGFVVDGAG-----FSGGAGNGGAGGAGANGGTGS 686
QY 536 CAGGGGGCAAGAGGAAATACGTGCTGT--GGGTCTCCCCCAACGCGCTCGCGGAT 593
Db 687 GNGGDDGG-AGGIGGAG-----GNGIPGTGTETPAGGT-----GAKGGD 724

```

OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248346; CAB95727.1; ..
 KW Complete proteome.
 SQ SEQUENCE 1938 AA; 152027 MW; 3CC73A507FEB0F9C CRC64;
 Query Match 14.2%; Score 820; DB 16; Length 1938;
 Best Local Similarity 32.7%; Pred. No. 3e-38;
 Matches 335; Conservative 14; Mismatches 475; Indels 201; Gaps 38;
 QY 1 GGGGCCCATCTCTTTCAGCCAGTCGCCAAG-----AATCATGAAGTCCCAAGTGCAG 55
 DB 692 GGTGCGAGAGGTGTGQADGNAGNGGDDGGVGGNGGNGADNTTAAAGTTGGAGGAGG 751
 QY 56 CACCGCCACCGCCCGCGGCCCCAGCTCGCGCTGAAGGC-----CGGCAA 103
 DB 752 TGGTGAAGTGTGCGQNGGN-----GGNG-GTGGKGTGAGWNSLDPLLAADGGQGG 805
 QY 104 GACAGCGAGCGGTGC-----GGCGAGGTGTGCGCTGTC-----TGTCTGACGAG 149
 DB 806 TGGTGNAGAGGTGTGQADGNAGNGDGGVGGNGGNGADNTTAAAGTTGGAGGAGGAG 865
 QY 150 AGCTGTGCCATCTCG-CCGTCGCGGGCGCGCGCGCGCTCCCTGCCCTGTGCACGA 208
 DB 866 GTGTTGGAAGTGTGCGQNGGNGGNGGTGKGTG-GDGLALSSGAGGKGGNGDAGK 924
 QY 209 GCAGCAGGTAACTGTCTCTACGATCAAGCGGTGTACTCAAGCTCAAGCGCTCAAGGCT 268
 DB 925 AGTGSAPGTAGTGDGKGK-----GNGGICAA--GTTGPGVGTGASGG--TGSAGGAGT 974
 QY 269 GGTGCCCACTGCCCGCCAGACCCCA--AGTGACAGAGTGGAGATTCTC-----CAGCAC 323
 DB 975 GGDG-----GAANGTAGAGAGNGKGGDGGAGVTSSTAGNSGGAGGG 1020
 QY 324 GTCATCGACTACAGGAGCCTTCAGTTGAGTGAAGTCACTCGGAATCCGAAGTTGGAGC 383
 DB 1021 GKGGDAGAGGAGATPGANGIAGN--GGDGGDGAAGAVGISGATGAGDGGHGGTGG---- 1073
 QY 384 CCCGGGGCCGAGG-----GCTGCCGCTCCGGCTCCGCTCAGCACCTCAACGGCGAGAT 439
 DB 1074 --AGNGGTGAGGSGGIDGVGGTGTGNGNGAIG-----GAGGDAGGSGN 1119
 QY 440 CAGCGCCTGACGCGGAGGCGCATGCGTTCCTCGCGGACGATCGCATCTGTGTGCTG 499
 DB 1120 SGNG--GTGKGGNAGAGGAG--SNGTGVANGTGD--GGNGAA--GAATAGSNG 1170
 QY 500 AAGCGCTCCCGCCAGGAGACCGCGGACCCAGCCATCCAGGGGC-----AAGAGGA 551
 DB 1171 GAGTG-----SAGNGGTGGRGS-----GGAGGDDGIGVGGKGGNGADGVEVGAGGA 1219
 QY 552 ATTACGTCTCTGTGGGTCTCCCGCAACGCGCTCCGCGGATCT--GAGCGAGAACAG 608
 DB 1220 GSGGNTSPGNGGGGCGGGGGAGGAGAG--GAGGNGTAGNGGCGGAGGTGGAG 1275
 QY 609 ACCGATCGCGGCCCATCTGCGCCCTTAATGTCATCCAGCTGG-----GAGG 650
 DB 1276 AASSATNGSGAGGTGGAG--GTGGAGDGVGAGGNGGHHGGDAGDGGNGANGNRS 1332
 QY 651 -----GGCTGAGCTGAGCACTGCGAGGAGGAGCGCTCTCTCTGACACCTACT 703
 DB 1333 SGSFLAAGGTGAAGDGGGGGGGAGG--GAGGCGGAGAG-----GTGNGGNTGCT 1384

QY 704 AGTCACACAGAGAC-----TTTAGGGGGTGGGATTCACACTCGTGTGTTCTTA 749
 DB 1385 AGTAGAAGNAGAGKAGGAGGGGGTGGGTGGGGAGGGGGTGGTGGTGGTGGTGGTGG 1444
 QY 750 TTTTGTGAAAACAGACATTTTAAAAAATGTCAC-----CGTTGGTGTCTTCACGATT 803
 DB 1445 QGKNAGGGGAGGGGAGGGGGGCGDGGDAGTGGGNGGNGNRNSNGTGGAGGN--GGGGANG 1502
 QY 804 TCTGAGGAAATG-----CTTGTATTGTATATACATGATCACCAGTCAAGATA 855
 DB 1503 GAGGAGGGGGTGGNGGAGGAGGAGGAGGNGTGGNGGNGG--IAGMGGNGGAGTGG 1561
 QY 856 TTGTTTT-----ACAATAGTCTGTGGGGCTGTTTTTTTGTATTAAACAAATAATTAGA 911
 DB 1562 NGGNGGGGNGGNGAGNGNSGTGSDGAGGNGAAGTGGTGGDGG--LTGTGTTGGSGG 1619
 QY 912 TGGTG 916
 DB 1620 TGGDG 1624
 RESULT 12
 OS3557 PRELIMINARY; PRT; 1079 AA.
 AC OS3557;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PGRS-family protein (Fragment).
 GN RV3512 OR MTU023.19.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogan A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL; AL022022; CAAL7749.1; ..
 DR PIR; B70807; B70807;
 DR Tuberculist; RV3512;
 DR GO; GO:0004420; F:hydroxymethylglutaryl-CoA reductase (NADPH)...; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR InterPro; IPR002202; HMG-CoA red.
 DR PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
 KW Complete proteome.
 FT NON_TER
 SQ SEQUENCE 1079 AA; 81163 MW; A79718CDBE74B97D CRC64;
 Query Match 14.2%; Score 819; DB 16; Length 1079;
 Best Local Similarity 31.9%; Pred. No. 2.2e-38;
 Matches 340; Conservative 16; Mismatches 462; Indels 249; Gaps 38;
 QY 2 GGGCCCATCTCTTTCAGCCAGTCGCCAAG-AATCATGAAGT--CGCCAGTGGCAGCAC 58
 DB 106 GTGSPAGTAGTGD--GGKGGNGGIGAGTTGTPGTGAGGTTGGGGAGTGGDGAAN 162
 QY 59 CGCCACCGCCCGCGCGGG-----CCCCAGTCCGCTGAAGCGCGGCAAGACAC 109
 DB 163 GGTAGAGAGGNGKGGDGGAGVTSSTAGNSG-GAGGSGGKGGDAGAGAGATPGANGTA 221

```
QY 314 T---CTCCAGCAGCTCATCGACTACATCATAGGACCTTCACTGGAGCTGAACCTCGAATC 370
Db 433 SGAGGTNGSGAGGT-----GGGGMGGAGGAGADNPTGTTGGTGGTGGTGAAGAGAGGA 488
QY 371 CGAAGTTG-----GGACCCCGGGGGCGAGGGCT-----GCCGGTCCGGCTCC 415
Db 489 AGTGGTGGMTGTTGNAGVGA-----GGGGGAGAGAGADADQPGATGTTGTFAGAGGA 543
QY 416 GCTCAGCACCTCAACGGCGAGATCAGCCCTGACGGCCGAGGGGG-CATCGTTCTT- 473
Db 544 G-----GAGGSSGAGTNGSGAGGTGGGAGGAGAGADNPTGTTGTTGGG 591
QY 474 GCGGACGA-TCGCATCTTGT-CTCGCTGAAGCCCTCCCGGAGGACCGGCGGA--CCCC 529
Db 592 GTGGAAGAGAGAGAGTGGTGGMTGTTGNAGV-----GAGGCGGAGGAGAGGGS 644
QY 530 AGCCATCCAGGGGC-----AAGAGGA-ATTAGTCTCTGTGGTCTCCCCCAACGGC 583
Db 645 SGAGGTNGSGAGGTGGGAGGAGAGADNPTGIG-----GTGGD----- 685
QY 584 CTCGCGGATCTGAGGAGAGAACAGACCGAT-----CGCGGCGCACTGCGCCCTTAAC 636
Db 686 ---GGTGAAGAGAGAGAGTGGTGGMTGTTGNAGVGGAGGGGGGAGAGADADQPGA 742
QY 637 TGCATCAGCCTGGGCTGAGGCTGAGGCTAGGCACTGGCGAGGAGAGGGCTCTCTTGCAC 696
Db 743 TG-CTGPAG--GAGGAGGAGGSSGAG--TNGSGAGGTGGGAG-----GAGI 787
QY 597 ACCTACTAGTACACAGAGACTTTAGGGGTGGGATCCACTCGTGTGTTCTATTATTG 756
Db 788 SFSNGSNGT-----GGTGG-----VGGTGGGAGNAGTAGDPPGKGTGGTGGSGAGSGG 839
QY 757 AAAAGCAGACATTTTAAATAATGGTCACTGTTGGTCTCTCAGATTTCTGAGGAATTG 816
Db 840 ANFNGTGGTGGTGGTGGGAGGAGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 899
QY 817 CTTTGTATTGATATTACATGATCAGGCTGAGATATTTTACATATGTTCTGTG 876
Db 900 PGSGGAGTGGTG-----GRGGAGGGGGAGGAGAT-GVGGGGG 940
QY 877 GGGCTGTTTTTGTATTAAACAAATAATTAGATGTTGAAATAAAA 924
Db 941 GNGGNGGNGTGVSGPGLGG-----AGTGGLGAGAGAGGA 977

RESULT 10
ID O17434 PRELIMINARY; PRT; 988 AA.
AC O17434;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
OS Minor ampullate silk protein Misp1 (fragment).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98200471; PubMed=9541398;
RA Colgin M.A., Lewis R.V.;
RT "Spider minor ampullate silk proteins contain new repetitive sequences
RT and highly conserved non-silk-like 'spacer regions'.";
RL Protein Sci. 7:667-672(1998).
DR EMBL; AF027735; AAC14589.1; --
FT NON_TER
SQ SEQUENCE 988 AA; 79082 MW; 461E03DF53F7085D CRC64;

Query Match 14.4%; Score 829.5; DB 5; Length 988;
Best Local Similarity 32.4%; Pred. No. 5.2e-39;
Matches 307; Conservative 16; Mismatches 363; Indels 261; Gaps 39;

QY 8 ATCTGTTTCAG-----CCAGTCGCGCAAGATCATGAAGTCGCCAGTGGCAGCACCGCC 62
```

```
Db 197 ASSVTTLTSGPAPVFPSPASGGYGAGAGAVAAAGAGGAGGAGGAGGAGGAGGAGGAG 256
QY 63 ACCCCCGCGCGCGCCCGCCAGCTGCGCGCTGAAGCCCGCAAGACAGGAGCGGTGCGG-- 120
Db 257 AGAGAAAAAGAGAG-----GAGGYGR-----AGAGAGAAAGAGAGAGAGAGGCGG 302
QY 121 -----GCGAGGTGGTGGCTCTCTCTCTGAGCAGAGAGCGTGGCCATCTCGCG 166
Db 303 GYGAGAGAGAGAGAGAGAGAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 358
QY 167 CTCGCGCGCGCGCGCGCGCGCGCTGCGCTGC-CCTGCTGGAGCAGCAGCAGCAGTAACTGC 225
Db 359 AAAAAAGAGSGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG- 416
QY 226 TGCTCTACGACATGAACCGCTGTTACTCAAGCTGAGGAGCTGGTCCCAACCTGCCCC 285
Db 417 -----AGAGGYGRGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 466
QY 286 AGAACCGCAAGGTGAGCAAGGTGAGATTTCTCAGCAGCTATCAGTATCATCAGGAGACC 345
Db 467 AGAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG- 518
QY 346 TTCAGTTGGAGCTGAACCTCG-GAATCCGAAGTTGGGACCCCGCGCGCGCGAGGCTGCCG 404
Db 519 -----GGAGGYGRGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG- 560
QY 405 GTCCGCGGCTCGCTCAGCACCTCAACGGCGAGATCAGCGCCCTGAGCGCGAGGCG-GC 463
Db 561 ---AGAGAAAAAGAGAGAG-----GGYGRGAG-AGAGYAGAGAGAGGAGGAGGAGAG 608
QY 464 ATGGCTTCTCGGACGATCGCATCTTGTGCTGCTGAAGCGCTCCCGCGCGAGGAGCGCG 523
Db 609 GAGAAAAAATGAGAGG-----YGRGAGAGAGAA-----AGAGAGTGGAG 648
QY 524 GACCCCGCATTCAGCGGGGCAAGAGAAATTAOTGCTCTGTGGGTCTCCCGCAACGCGC 583
Db 649 -----YGGGGYGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG- 685
QY 584 CTCGCGCGGATCTGAGGGGAGACAGACCGATCGCGCGCACTGCGCCCTTAACCTGATCC 643
Db 686 -----AAAGAGAGAAAGAGAGAG--YGGGGYGAGAGAGAGAGAGAGAGAGAGAG- 730
QY 644 AGCCTGG--GGCTGAGGCTGAGGCACTGGCGA-GGAGAGGGGCGCTCTCTCTGCAACCTA 701
Db 731 AGYSRGRAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG- 772
QY 702 CTAGTCAACGAGACTTTAGGGGGTGGGATCCCTCGTGTGTTTCTATTATTTTGAAGA 761
Db 773 -----AAAAAGAG-----SGAGGYGRGAGAGAGAGAGAGAGAGAGAGAGAG- 811
QY 762 CAGACATTTTAAAAAATGGTCACTGTTGTTGTTCTCTCAGATTTCTGAGAGAAATTCCTTG 821
Db 812 YGGGGYGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG- 851
QY 822 TATTGTATTATTAATGATCACCAGTGAATATTTTACATAGTTCTGTGGGG-- 879
Db 852 -----YGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG- 872
QY 880 -----CTGTTTGTGTTTATTAACAAATAATTTAGATGGTGA 917
Db 873 DKEIACWSRCRTVASTTSLSSAEASSISSAASSTLVGGYLNTAA 919

RESULT 11
QY7TWC0
ID Q7TWC0 PRELIMINARY; PRT; 1938 AA.
AC Q7TWC0;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE PE-PGRS family protein.
GN PE_PGRS55 OR MB3541.
```



```

DE PE PGSR family protein.
GN MT3615.3.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteriales; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA DeCher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007164; AAK47974.1; -.
DR TIGR; MT3615; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR00084; PE_region.
DR InterPro; IPR002173; PfKB.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.
DR ProDom; PD01223; PE_region; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 3.
DR NCBI_TaxID=1773;
SQ SEQUENCE 1217 AA; 94463 MW; 2F5DDCA5A30A056E CRC64;

Query Match 15.4%; Score 889.5; DB 16; Length 1217;
Best Local Similarity 34.7%; Pred. No. 2.9e-42;
Matches 350; Conservative 20; Mismatches 471; Indels 167; Gaps 41;

QY 1 GGGGCCCATCTCTTTTTCAGCCATCGCCAGATCATGAAGTCGCGAGTGGCGAGCACCG 60
DB 180 GAGGWLFEVGGAGVGGAG--GGTGGAGGPGGLIWGGGGAGVGGAGGTGGAGGGA--E 235
QY 61 CCACCGCGCGCGCGCGCCAGCTGCGCTGAAGCGCGGCAAGACAGCG--AGCGGTGC 118
DB 236 LFFGAGGAGAGTDDG---PGATG-GTGCHGGVGGDGGCLAFGAGGAGCGGAGGAGS 290
QY 119 GGGC--GAGGTGCT---GGCTGTC-----TGTCTGACAGAGGTGCCATCTCGG 166
DB 291 DGGALGGTGTGTGGAGGAGGAGGALLLGGAGGQGGG-LGGAGGQGGMGAGGAGADNPTGTG 350
QY 167 CTGCGCGGCGCGCGCGCGCGCTG---CCTGCTCTCTGTCGACGAGCAGAGTAA-ACG 222
DB 351 GTG--GDCGTGSAGG- GAGGAAGQLFSAGAGNAGVGGAGGQGGGAGGAGADADQ 407
QY 223 TGCTGCTTACGATGAACG--CTGTACTACGCTCAAGAGCT-----GCTGCC 274
DB 408 PGATGTTGTFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 465
QY 275 CACCTGCGCCAGAACCGCAAGGTGAGCAA--GGTGGAGATTCTCCAGCAGCTCATCGAC 332
DB 466 GADNPTGTGGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 523
QY 333 TACATCAGGAGCCTTCACTTGGAGCTGAACCTCGGAATCCGAAGTTGGGACCCCGGGGG 392
DB 524 -----AGGAGADADPGATGTTGTFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 573
QY 393 CGAGGCTGCGCTCGCGCTCCGCTCAGCACCCCTCAACGCGCAGATCAGCGCCCTGACG 452
DB 574 CGAGGSSGAGTNGSAGGTGGGAGGA-----GGAGADNPTGTGTGGTGGTGGT 625
QY 453 GCCGAGCGGCA-----TG--CGTTCTCTCGGA-----CGATCGCATTTGTGTCTGAA 501
DB 626 GAAAGAGGAGGAAGTGTGTTGTTGNAGVGGAGGQGGDGGAGGAGADADPGATGTTGTA 685
QY 502 GCGCTTCCCGAGGAGCGGGGACCCAGCCATCCAGGGGCG-----AAGAGGA-ATTA 555
DB 686 G-----GAGGAGGAGGAGGAGN-----SGAGTNGSAGGTGGGAGGAGGAGGAGDNP 732
QY 556 CGTGTCTGTGGTCTCTCCCAACGCGCTCGCGGATCTCGAGGAGAGCAAGACCGAT- 614

```

```

DB 733 TGIG---GTGGD-----GTGGAAGAGGAGGAGTGGTGGMIGTTG 770
QY 615 -----CGCGGCGCACTGCGCCCTTAACCTGATCCAGCTGGGCGTGGAGCTGAGGCAC 668
DB 771 NAGVGGAGGCGGAGGAGGAGADADPGATG-GTGFAAGAGGAGGAGGAGGAGGAGTNGS 829
QY 669 GGC-GAGGAGAGGCGCTCTCTGACACACTTACTAGTCCACGAGACTTTAG--GGGG 725
DB 830 GGAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 883
QY 726 TGGGATTCACCTCGTGTGTTCTTATTTTGAAGAGCAGACATTTTAAAAAATGGTCA 785
DB 884 TGGTGTGGSGAG-GSGGANFNGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 942
QY 786 TTTGGTCTCTCTAGATTTCTGAGGAATTCCTTTGTTATTGTA-----TATTA 833
DB 943 -GKGTGAGDSDSAGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1000
QY 834 CAATGATCACCGACTGAGAA--TATCTTTTACATAGTTCTGTGGGCTGTT----- 884
DB 1001 GSGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1056
QY 885 -TTTTGTTATTAAACAAATAA-----TTTATGATGTTGAAAAAAA 924
DB 1057 GTGGTGGVGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1104

RESULT 5
Q964F4 PRELIMINARY; PRT; 2655 AA.
AC Q964F4
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fibroin.
OS Antheraea yamamai (Japanese oak silkworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Saturniidae; Saturniinae; Saturniini; Antheraea.
OX NCBI_TaxID=7121;
RN [1]
RN SEQUENCE FROM N.A.
RC TRANSPOSOME=mariner-like element;
RA Lee J.-S., Hwang J.-S., Kang S.-W., Suh D.-S., Jin B.-R., Kim Y.-S.,
RA Lee K.-S., Goo T.-W., Yun E.-Y., Kim K.-Y.;
RT "Complete nucleotide sequence of fibroin gene cloned from Antheraea
RT yamamai.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325500; AAK83145.1; -.
SQ SEQUENCE 2655 AA; 218672 MW; 89BB810EF9CC4CA6 CRC64;

Query Match 15.2%; Score 879; DB 5; Length 2655;
Best Local Similarity 26.7%; Pred. No. 2e-41;
Matches 315; Conservative 57; Mismatches 531; Indels 278; Gaps 33;

QY 1 GGGGCCCATCTCTGTTTCAGCCAGTCTG-----CCAGAAATCATGAAAGT 43
DB 1194 GYGVSDSAAAAAASGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1253
QY 44 CGCCAGTGGCAGCACCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103
DB 1254 RDGGYGVSGSSAAAAAASGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1305
QY 104 GACACGAGCGGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 156
DB 1306 AAAAAAGSAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1364
QY 157 ----CCATCTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 209
DB 1365 AAAAAAAGSAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1424
QY 210 CAGCAGGTTAAACGTCTGCTG-----TCTACGACATGAAACGCTGT-----TA 250

```



```

QY 61 CCACCGCGCGCGG--GCCCGAGCTGGCGCTGAAG--GCCGCGACAGACGAGCGGT 116
Db 1033 STGATGATGATGATGATGATGATGATG--STGATGATGATGATGATGATGATGATG 1091
QY 117 GCGGCGGA--GGTGGG--CGCTGTCTGTCTGACGACGCGTGGCCATCTCG--CGCTCCCGG 173
Db 1092 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1150
QY 174 -GGCGCGCG--GGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
Db 1151 TGSTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1210
QY 232 ACGACATGAACGGCTGTACTACAGCTTCAAGGAGCTGTGGCCACCGCTGCCCGAACCC 291
Db 1211 TGSTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1267
QY 292 GCAAGTGAGCAAGGTGAGGATCTCCAGACGCTCATCGACTACATCAGGAGCTTCACT 351
Db 1268 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1327
QY 352 TGGAGCTGAATCG-----GAATCGAAGTTGGGACCCCG--GGGCGCGAGG--CGTGC 402
Db 1328 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1387
QY 403 CGGTCCGGCTCGCTCAGACCCCTCAACGGCGA--GATCAGCGCCCTGACG--GCCGAGG 459
Db 1388 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1443
QY 460 CGCATGCGCTTCTCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 519
Db 1444 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1502
QY 520 GGCG--GACCCAGCCATCCAGGGGCG--AAGAGGAATTAAGTGT--CTGTGGCTTCCCG 575
Db 1503 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1558
QY 576 CAACGCGCTCGCGATCTGAGGAGCAACAGCCGATCGGCGGCGGCGGCGGCGGCGGCG 635
Db 1559 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1614
QY 636 CTGCATCAGCGCTGGGCTGAGGCTGA--GGCAGCTGGCGA--GGAGAGGCGGCTCTCTCTG 693
Db 1615 ATG--GTGATG--GTGATGAGGAGGATGATGATGATGATGATGATGATGATGATG 1672
QY 694 CACACTACTAGTAC--CAGAGACTTGGGGGTTGG--GATTCACCTCGTG--TGTTCTAT 750
Db 1673 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1732
QY 751 TTTTGAAGCAGACATTTTAAAGATGTCACGTT--TGCTGCTTCTCAGATTTCTGA 808
Db 1733 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1792
QY 809 GGAATTGCT--TTGTATGTATATTACATGATCAGGACTGAGATATTGTT-----T 861
Db 1793 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1847
QY 862 TCAATAGTCTG--TGGGGCTGTT--TTTTTGTATTAAACAAATTAATTAGTGGTCAA 919
Db 1848 GSTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1905
QY 920 AAAAAA 926
Db 1906 STGATGA 1912

```

RESULT 2

```

Q98MG7 PRELIMINARY; PRT: 3145 AA.
AC Q98MG7
DT 01-OCT-2001 (T-EMBLrel. 18, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical glycine-rich protein mlr0587.

```

```

GN MLR0587.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
RN NCBI_TaxID=381;
[1]
SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpko S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002995; BAB48146.1; -.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 38
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 3145 AA; 271616 MW; F10C498282766FB CRC64;

Query Match 17.8%; Score 1028.5; DB 16; Length 3145;
Best Local Similarity 37.2%; Pred. No. 1.1e-49;
Matches 369; Conservative 18; Mismatches 499; Indels 107; Gaps 41;

QY 1 GGGGCCCATCTCTTTCAGCCAGTCCGACGATCATGA-----AAGTCGCCAGTGGCAG 55
Db 1251 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1309
QY 56 CACCGCCAGCCCGCCCGCG--GCCCGAGCTG--CGGCTGAAG--GCCGCGAAGACAGC 109
Db 1310 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1369
QY 110 GAGCGGTGGCGG--AGGTGGTGC--CCTGTC--TGCTGAGCAGAGCGTGGCCATCTCGC 165
Db 1370 TGSTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1429
QY 166 -GCTGCCG--GGCGCGCG--GGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 219
Db 1430 TGSTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1489
QY 220 ACGTGTCTCTACGACATGACCGCTGTACTCAGC-----CTCAAGAGCT 268
Db 1490 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1547
QY 269 GGTGCCACCTCCCGCCAGAACCGGAGGTGAGGAGGTGGAGATCTCCAGCAGCTCAT 328
Db 1548 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1607
QY 329 CGACTACATCAGGAGCTTCAGTTGAGCTGAACTCG--GAATCCGAACTTGGGACCCCC 386
Db 1608 GSTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1660
QY 387 GGGGCGCGAGG--GCTGCGCGTCCGGCTCCGCTCAGCAGCCCTCA--ACGCGAGATCAG 442
Db 1661 TGDTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1720
QY 443 CGCCCTGACG-----GCCGAGGCGGATCGTTCCTTCGCGAGCGATCGCATCTTGTGTGCG 497
Db 1721 TG--ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1777
QY 498 TGAAGCGCCCTCCCGAGGACCGCGGACCCAGCCATCCAGGGGCG--AAGAGGATAC 556
Db 1778 TGATG-----ATGDTGATGATGATGATGATGATGATGATGATGATGATGATG 1827
QY 557 GTGCT-CTGTGGGTCTCCCGAACCGCGCTCGCGGATCTGAGGGAGAGAACAGCCGATC 615
Db 1828 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1882
QY 616 GCGCGCCAGTGC-----GCCCTTAACATGCATCCAGCCCTGGGCTGAGGCTGA--GGCAGCTGG 670

```


Db 261 CPGNCK--NGGACVGVNTYNCRCPEVTGQVCTEDVDEQCLMFNAQAGTCHNTHG 318
Qy 217 TAAACGTGCTCT--ACGACATGAACGGCTGTACTC-----AC-- 255
Db 319 YNCVUNGTGDCSENIDDCASACFQATGATCHDRVAFYCEPHGRTGLLCHLKHACIS 378
Qy 256 -----GC-----CTCAAG--GAGCTGGTGCC-----CACCTG-CCCAGAAC 291
Db 379 NPCNEGNCDTNPVNGKRICTCPSGYTPACSDQVDECDLGNRCEHAGKCLNTLGSFEC 438
Qy 292 GCAAGGTGAGCA-----AGGTGGAGATTCTCCAG-----CAGGT 325
Db 439 QCLQYTGCGCEIDVNECISNFCNDATCLDQIGFQICMPGVEGYCEINTDECASSP 498
Qy 326 CATCGAC-----TACAT--CAGGGACCT-----TCAGTTGG 354
Db 499 CLFHGCMKDIHEFOCCPKGFNGHLCOYDVDECASTPCXGAKCLDGPNTYTCVCTEGY 558
Qy 355 AG-----CTGAACTCGAA--TCCGAAGTTG-----GGAC 382
Db 559 TGTCEVDIDCDPDPCHYGSCKDGVAFITCLCPGYTGHICEFTNINECHSOPCRHGTC 618
Qy 383 -----CCCCGG-----GGGC-----CGAGGGCTG--C-- 402
Db 619 QDRDNSYLCILKLTGTPNCEINLDDCASNPDCSGTCLDKIDGYECACEPGYTGSMCNVN 678
Qy 403 ---CGGTCCGGGCTC-----CGC-----TCAGCACCTCAAC--GGCGAGATCAG 442
Db 679 IDECAGSPCHNGTCDGDIAGTCTCPGYHDPITCLSEVNECNSNPCLIHGACRDGLN--G 736
Qy 443 CGCCCTGACGG-----CCGAGCGGCGATCGTTCTCTGGG----- 477
Db 737 YKDCAPGWSGTNCDINNNECESNPCVNGGTCKMTSGYVCTCREGFSGPNQININECA 796
Qy 478 -----ACGATCG-----ATCTTGTGCTGCTGAAGGCG-----TCCCC 511
Db 797 SNPLNQGTICIDDVAGYKNCPLPYTGATCEVVLAPCATSPCKNSGVCKESEDYESFSCV 856
Qy 512 CAGG-----GACCGCGGA--CCCCAG-----CCATCC 537
Db 857 CPTGWOQTCEVDINECVKSPCRHGASCONTNGSYRCLCPGFTGSCYQYDVNECDSPCLHGTC 916
Qy 538 AGGGG-----CAAGAGGA-----ATTAGTGTCTGT----- 565
Db 917 HNGSCTDGINTAFCDCLPFGQAFCEEDINECASNPQNGANCTDCVDSYTCPCVGFN 976
Qy 566 -----GGGTC-----TCCCCAACCGCG-----C-TGCGCGGATC 594
Db 977 GIHCENNTPDCTBSSCFNGGTCVDGINSFTCLCPGFTGSCYQYDVNECDSPCLHGTC 1036
Qy 595 TGAGGGAGAACAGA-----CCGATCGCGGC-----CAC-----TGCGCCC 631
Db 1037 QDSYGTCTCPOGYTGLNQNLRVNCDSAPCKNGRCRWQNTQVHCERSGWTGVNCDV 1096
Qy 632 TTAACTGCA-----TCAGGCTGGGGTGAG-----GCTGA-----G 663
Db 1097 LSVSCEVAQXRGIDVTLICQHGGLCVDEGDKHYCHYQAGYTGSCYCEDEVNDECSNPQN 1156
Qy 664 GCACT--GG--CGAGGAGA-----GGGC-----GCTC----- 686
Db 1157 GATCTDYLGGFCKCVAGYHGSNCSEBINECLSQPCQNGGTCTDITNSYKSCSPRGTOGV 1216
Qy 687 -CTCTCTGC-----ACACTACTAGTACACAGACTTTAG-GGGTGGGATTCACCTC 738
Db 1217 HCEINVDCHPPLDPAASRPFKNNGTCVDQVGGYTCCTPFGVGERCEGDVNECLSNPC 1276
Qy 739 GTGTGTTTCTATTTTGAAGAAGCAG-ACATTTTAAAA--AATGTCACGTTTGGTGCTT 795
Db 1277 DP-RGTQNCVQRVNDHFCECRAGHTGRCESVINGCRGPKCKNGGVCAVASNTARGFICR 1335
Qy 796 CTGAGATTCTGAGGAATGCTTTGTTATGATATTAC--AATGATC-----ACGA-- 846

Db 1336 CPAGFEGATCENDARTCGSLRCLNGGTCTSGRSPRTCLCLGSPGTGPECQFPASSPCVGSN 1395
Qy 847 -CTGAGAATATGTTTACAATA---GTTG-----TGTGGGG 879
Db 1396 PCYNGTCEPTSENPFYKCLCPAKFNGLLCHILDYSFTGGAG 1437

Search completed: May 7, 2004, 15:18:20
Job time : 32 secs

[9] POST-TRANSLATIONAL PROCESSING.
RP MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki I., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN [10]
RP INTERACTION WITH DTX1 AND DTX2.
RX MEDLINE=21123790; PubMed=1226752;
RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
RA Okano H., Matsuno K.;
RT "Murine homologs of *deltex* define a novel gene family involved in
vertebrate Notch signaling and neurogenesis.";
RL Int. J. Dev. Neurosci. 19:21-35(2001).
RN [11]
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
Upon ligand activation through the released notch intracellular
domain (NICD) it forms a transcriptional activator complex with
RBP-J kappa and activates genes of the enhancer of split locus.
Affects the implementation of differentiation, proliferation and
apoptotic programs (By similarity). May play an essential role in
postimplantation development. Probably in some aspect of cell
specification and/or differentiation. May be involved in mesoderm
development, somitogenesis and neurogenesis. Involved in the
maturation of both CD4+ and CD8+ cells in the thymus.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
terminal fragment N(EC) which are probably linked by disulfide
bonds. Interacts with DTX1 and DTX2.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
proteolytical processing NICD is translocated to the nucleus.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=Q01705-1; Sequence=Displayed;
Name=2;
IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
thymus. Expressed at lower levels in the spleen, bone-marrow,
spinal cord, eyes, mammary gland, liver, intestine, skeletal
muscle, kidney and heart.
CC -!- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
endothelial cells, while much lower levels are seen in the
neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
the neuroepithelium. At 13.5 dpc expressed in the surface
ectoderm, eye and developing whisker follicles.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
which is proteolytically cleaved by a furin-like convertase in the
trans-Golgi network before it reaches the plasma membrane to yield
an active, ligand-accessible form. Cleavage results in a C-
terminal fragment N(TM) and a N-terminal fragment N(EC). Following
ligand binding, it is cleaved by TNF-alpha converting enzyme
(TACE) to yield a membrane-associated intermediate fragment called
notch extracellular truncation (NEXT). This fragment is then
cleaved by presenilin dependent gamma-secretase to release a
notch-derived peptide containing the intracellular domain (NICD)
from the membrane.
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 36 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; Z11886; CAA77941.1; -
DR EMBL; L02613; AAK14898.1; -
DR EMBL; X68278; CAA48339.1; -
DR EMBL; AJ238029; CAA40733.1; -
DR EMBL; X82562; CAA57909.1; -
DR FIR; A46019; A46019.
DR PIR; B49175; B49175.
DR HSSP; P00740; LEDM.
DR MGD; MGI:97363; Notchl.
DR GO; GO:0005887; C:integral to plasma membrane; IC.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0030154; P:cell differentiation; IMP.
DR GO; GO:0007386; P:compartment specification; IMP.
DR GO; GO:0007219; P:N signaling pathway; IC.
DR GO; GO:0045944; P:positive regulation of transcription from P. .; IDA.
DR InterPro; IPR001152; ASX hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR008000; Notch_dom.
DR Pfam; PF00023; ank; 7.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00666; Notch; 3.
DR FIRSF; FIRSF002279; Notch; 1.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00004; NL_2.
DR PROSITE; PS00297; ANK_REPEAT; 1.
DR PROSITE; PS00088; ANK_REPEAT; 2.
DR PROSITE; PS00100; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS00026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Receptor; transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 2531
FT CHAIN 1711 2531
FT CHAIN 1744 2531
FT CHAIN 19 1755
FT DOMAIN
Query Match 9.6%; Score 553; DB 1; Length 2531;
Best Local Similarity 22.0%; Pred.No. 5.3e-21;
Matches 313; Conservative 38; Mismatches 493; Indels 578; Gaps 78;
QY 28 CAGATCATGAAGTCGCGCAGGAG-----CACCGCC-----62
DB 24 CSOPSGTCLNG---GRCEVASGTETACVAGSFVGRQCQDPNCLSTRCKNAGTCYVDHG 80
QY 63 ----ACCGC-----CGCCGC-GGGCCCGAGTCGCGCT-GAAGGCC---98
DB 81 GIVDYACSCPLGSPGLCLTLPDKLANPCANGTCDLLTLEYKRCSPGWSKSCQ 140
QY 99 -GCCAGACAG------CGAGCGTGC-----GGCGAGG-----TG 128
DB 141 ADPCANPCANGCCCLPFSSVYTCRPPGHGTCRQDVNCSQNFCLRHGHCHNEIG 200
QY 129 GTGGCTGTCTG-----TCTGAGCA-GAGC-GTGGCCATCTCGGCTG--C-----170
DB 201 SYRCACCAHTGHCELPYVPCSPQNGATCRPTGDTTTCACLPFGAGQCNENVD 260
QY 171 CGGGGCGCCGGGGCGCGCTGCC-----TGCCCT-----GCTGGAC-GAG-CAGCAGG 216

Db 438 CQCIGYTGPRCEIDVNEICSNPCQNDATCLDQIGFQICMPGEGVYCBINTDEBASS 497
QY 325 TCATCGACT-----ACAT--CAGGCACT-----TCAGTTG 353
Db 498 PCLHNGRCVCKINEFLCQPKFGSHLCQYDVDECASTCKNGAKCLDGNNTYTCVTEG 557
QY 354 GAG-----CTGAACGTG---GAA--TCCGAGTTG-----GGA 381
Db 558 YTGTHCEVDIDEDDPCDHGLCKDGVATFTCLCQPGYTGHHCBTWINECHSQPCRHGT 617
QY 382 C-----CCCCG-----GGC-----CGAGGGCTG--C-- 402
Db 618 CQDRNYYLCLCLAGTTPNCEINLDDCASNPDCSGTCLDKIDGVECACEPGYTGSMCV 677
QY 403 ---CGGTCCGGGCTC-----CGC-----TCAGCACCTCAAC--GGCGAGATCA 441
Db 678 NIDECAGSPCHNGTCEGDIAGFTCRPEGYHDPDTCLSEVNECNSPCLHAGCRGLN-- 735
QY 442 GGGCCCTGACCG-----CGAGGGGGGANGGTTCTCTCGCG----- 477
Db 736 GYKCDCAFGWGTNCDINNNECESNPVNGTCKDMTSGYVCTCREGFGSPNCOTNINBC 795
QY 478 ---ACGATGCG-----ATCTTGTCGTGCTGAAGCGCC-----TCCC 510
Db 796 ASNPCLNQTGIDDVAGYKNCPLPYTGATCEVVLPATSPCKNSGVCKSEDEYFSC 855
QY 511 CCAGG-----GACCGGGGA--CCCCAGCCATCCAG----- 539
Db 856 VCPGTGOGTCEIDINECVKSPRHHGASCONTRGSRVCLQAGYTGRENCESDIDCRPNP 915
QY 540 ---GGGGCAAGAGGA-----ATTACGTGCTCTG-----TG- 566
Db 916 CHNGSCTDGVNAFAFCDCLPQFQAFCEEDINECATNPQNGANCCTCDVDSYTCCTPTGF 975
QY 567 -----GGTC-----TCCCCCAACGGC-----C-TGCGCGAT 593
Db 976 NGIHCENNTPTCESSCFNGGTGVDGINSFTCLCPGFTGSGYQYDVNEDSRPCLHGT 1035
QY 594 CTGAGGAGAAACAAG-----CCGATCGCGCG-----CAC-----TGCGCC 630
Db 1036 QDSVGTGKTCPOGYTGLNQNLRVNCDSAPCKNGKCKWQNTQYHCEBSGWTGNCND 1095
QY 631 CTTAACTGCA-----TCCAGCTGGGGCTG-----AGGCTGA----- 662
Db 1096 VLSVCSVAARXRGIDVTLLQHGGLCVDEBKHVCHCQAGYTGSCYCEBDEVBCSPNPQ 1155
QY 663 GGCACCT--GG-----CGAGGAGAG-----GGC-----GCTCCTCTCT--GCACACCTA- 701
Db 1156 NGATCTDYLGGFCKVAGYHGSNCBBIINECLSQPCQNGGTCTDILNTYKCSCPRTGQ 1215
QY 702 ---CTAGTCAC-----CAGACACTTAG-----GGTGGGATTCCTACT 737
Db 1216 VHCENVDCHPPLDPAKSPKCFNNGTVDQVGGYTCPPGFVGERCEGVNECLSNP 1275
QY 738 CGTGTGTTTCTATTTTGAAGAGCAG--ACATTTTAA--AATGTCACGTTTGTGCT 794
Db 1276 CDP--RTGNCQVRVNDHCECRAGHTGRCEVINGCRKPCRGVCAVASNTARGFTC 1334
QY 795 TCTGATTTCTGAGAAATGCTTTGATGATATATAC--AATGATC-----ACCA 846
Db 1335 RCPARFEGATENDARTCGSLRCLNGGTGTCISGPRSPCLCLGSGFTGPEQCPASSPCVGS 1394
QY 847 ---CTGAGATATGTTTACAATA---CTTC-----TGTGG 877
Db 1395 NPYNGTCEPTSESFFVRLCPAKFNGLLCHLDYSFTGAAG 1437

RESULT 15
NTCL MOUSE
ID NTCL MOUSE
AC Q01705; Q06007; Q91905; Q93JCC; Q9QW58; Q9RCX7;
DT 01-NOV-1995 (Rel. 32, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Motch A)
DE (MT14) [p300].
GN NOTCH1 OR MOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Embryo;
RX MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA Copeland N.G., Gridley T.;
RT "cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homolog of Drosophila Notch.";
RL Genomics 15:259-264(1993).
RN [2]
RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RC STRAIN=CD-1; TISSUE=Embryo;
RX MEDLINE=93050801; PubMed=1426644;
RA Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
RT "Expression analysis of a Notch homologue in the mouse embryo.";
RL Dev. Biol. 154:377-387(1992).
RN [3]
RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RC TISSUE=Embryo;
RX MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Notch, a mouse homologue of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development.";
RL Development 115:737-744(1992).
RN [4]
RP SEQUENCE OF 1161-1547 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [5]
RP SEQUENCE OF 1659-1673 FROM N.A.
RX MEDLINE=99364499; PubMed=10437788;
RA Lee J.S., Ishimoto A., Yanagawa S.;
RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
RL FEBS Lett. 455:276-280(1999).
RN [6]
RP SEQUENCE OF 1950-2201 FROM N.A.
RX MEDLINE=98029496; PubMed=9384671;
RA Messerle M., Folio M., Nehls M., Eggert H., Boehm T.;
RT "Dynamic changes in gene expression during in vitro differentiation of
RT mouse embryonic stem cells.";
RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN [7]
RP SEQUENCE OF 1655-1659. CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
RX METAGENESIS OF 1651-ARG-ARG-1654.
RX MEDLINE=98318619; PubMed=9653148;
RA Logeat F., Bessia C., Brou C., Lebail O., Jarriault S., Seidah N.G.,
RA Israel A.;
RT "The Notch1 receptor is cleaved constitutively by a furin-like
RT convertase.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
RN [8]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).

EGF-LIKE 29. CALCIUM-BINDING (POTENTIAL).
EGF-LIKE 30. CALCIUM-BINDING (POTENTIAL).
EGF-LIKE 31. CALCIUM-BINDING (POTENTIAL).
EGF-LIKE 32. CALCIUM-BINDING (POTENTIAL).
EGF-LIKE 33. CALCIUM-BINDING (POTENTIAL).

T	DOMAIN	1097	1141
T	DOMAIN	1145	1181
T	DOMAIN	1191	1219
T	DOMAIN	1221	1265
T	DOMAIN	1267	1305
T	DOMAIN	1307	1346
T	DOMAIN	1348	1384
T	DOMAIN	1387	1426
T	REPEAT	1445	1480
T	REPEAT	1481	1522
T	REPEAT	1523	1562
T	REPEAT	1517	1546
T	REPEAT	1580	1590
T	REPEAT	1584	1613
T	REPEAT	2017	2046
T	REPEAT	2050	2079
T	DOMAIN	1730	1733
T	DOMAIN	1891	1894
T	DOMAIN	2258	2261
T	DOMAIN	2497	2500
T	SITE	1654	1655
FT	DISULFID	24	37
FT	DISULFID	31	46
FT	DISULFID	48	57
FT	DISULFID	63	74
FT	DISULFID	68	87
FT	DISULFID	89	98
FT	DISULFID	106	117
FT	DISULFID	111	127
FT	DISULFID	129	138
FT	DISULFID	144	155
FT	DISULFID	149	164
FT	DISULFID	156	175
FT	DISULFID	182	195
FT	DISULFID	199	204
FT	DISULFID	206	215
FT	DISULFID	222	233
FT	DISULFID	227	243

Query Match
 Best Local Similarity 21
 Matches 309; Conservativ

Query Match	9.6%;	Score 554;	DB 1;	Length 2531;
Best Local Similarity	21.7%;	Pred. No. 4.7e-21;		
Matches 309;	Conservative 33;	Mismatches 499;	Indels 582;	Gaps 75;
28	CAAGAAATCATGAAAGTCGCGAGTGGCAGACACGCG			68
Db	-----GRCEVANGTEACVCSGAFVQRCQDPSPCLSTPFCRNAGTCYVV			77
24	CSQPSGTCING-----GCCGC			69
Qy	-----GGCCCCAGCTGGCGCG-TCGAAGC			97
Db	78 DHGGIVDYACSCPLGFGSPGLCTPLANACLNPCRNGTCDLLLTLEYKRCPCPPGNSGKS			137
Qy	98 C---GGCAAGACA-----GGCAGCGGTGC-----GGCGAGGT-----			127
Db	138 CQQADPCASNPCANGGQCLPFESSY:CGCPPGPHGPTCRQDVNECSQNPLGRHGGTCHN			197
Qy	128 ---GGTGCCTGTCTG---TCTGAGCAGAGCGTGCGC---ATCTCGCGCTGCGCGGC---			176
Db	198 EIGSYRCACRATHGPHCELPYVPCSPFSPQNGGTCRPTGDTIHCACILPFGAQNCEN			257
Qy	177 -----GCCGGCGCGCTGCG-----TGCGCT-----GCTGGAC-GAG-CAGCAG			215
Db	258 VDCPFGNCKNGGACVDGVNTYNCRCPPPEWTGQYCTEDVDECOLMPENACQNAGTCHNSHG			317
Qy	216 GTAACGTCGTCTCT-----ACGACATGAACGGCTGTACTC-----AC-			255
Db	318 GYNCVGVNGWTGEDCSNIDDCASAAFCQGATCHDRVASFYCECPHGRTCGLLCHLNDA			377
Qy	256 -----GCGTCAAG---GAGCTGGTGCCC-----ACCCTG-CCCCGAGAC			290
Db	378 SNPCNEGSNCDTNPVNGKAICTCPRGYTGPAQSDVDEGALGANPCEHAGKCLNTIGSFE			437
Qy	291 CGCAAGGTGAGCA-----AGGTGGAGATTCTCCAG-----CACG			324


```
Db 628 SNPTFGVCRDINRYDCVQCPGFTGPICNVEINEECASSPCGEGSCVTCGSEGHFCLCPP 687
Qy 446 -----CTTGAC-----GG-----CGAGGCG-GCATGGTTCCTG-----C 475
Db 688 GSLPELCLPANHPCAHKCSHGVCCHDAFGPGFCQVCDPWSGPRCSQSLAPDACESQPCQA 747
Qy 476 GGAC-----GATCGCA-----TCTTGTCG- 496
Db 748 GGTCTSDGIGPHCTCAPGFGQHCHEVLSPTSLCEHGHCHESDPDQLTVCSPPGWGP 807
Qy 497 -----CTGAGC-----GCCT-----CC-----C-----CAGGAC 519
Db 808 RQQDVDEACAGSPGPHGTCTNLPSPRCHGTYGTFCDQIDCDPNCLNGSCQ 867
Qy 520 GCGGACCCACCCATCCAGGCGGCAAGAGGAATACGTGCTCTG-TGGGTCTCCCCAA 578
Db 868 DGVGFSFSCLSGPA-----GPRCARDVDECLSFSCGPG-TCTDHVASFTICFPGYGG 920
Qy 579 CGC-----GCTCGCGCGATCTGAGGGAGACA---AGACGATCGCG-----GCCA 623
Db 921 FHCETDLDCSPSSCFNGTCTVDGVSFSCLCRPGYTGHCQYKVDPCFRPCLHGGICN 980
Qy 624 CTGCGCCCT-TAACTG--CAT-----CCAGCCTGGGCTGAGG----- 658
Db 981 PTHSGPECTCRBFTGNOCNPVDWCSQAPCQNGRCVQTGAYCICPPENSGPLCDIPSL 1040
Qy 659 -CTGAG--GCACTGGCAGG-----AGAGGCG-----GCTCCTC--T 690
Db 1041 PCTEAAHGVLEOLCOAGGQCIDKDHSHYCVCPGMRGSHCHQVDPCTPAQCQHGTT 1100
Qy 691 CT-----GCACACTACTAGTC-----AC-----CAGAGACTTTAGGGGTGGATTCACCT 737
Db 1101 CRGYMGVYCEPTGYSDSCDEDDVDECAQSPQCGGSCIDLK-----AHYLCSP 1151
Qy 738 CQT-----GTGTT-----TC--TATTTTGAAGAAGC----- 762
Db 1152 PGTGLVLCBINEDDCGPGLDGLRGHLCVLDVGLGFCNCPGPGVTLGHCEADINEC 1211
Qy 763 -AGACATTTTAAAAATGTCACGTTGTTG--CTTCTCAGATTCTGAGGAATTCCTT 819
Db 1212 RPTCHAAHTRDCLQDPGHPRCICLPFTGPRQATLPCEQSCQHGQCRPSLG-- 1268
Qy 820 TGTATTGTATATAC-----AATGATCAGCATGA 850
Db 1269 ----RGGLTFTCHCVQPPWGLURCERVARSRELQCPVIGPCQQTARGPCACPPGLSGP 1324
Qy 851 G---AATATTGTTTACATA---GTTT-----TGTTGGGC-TGTTTTTT 889
Db 1325 SCRVSRAPSGATNTSCATPCLHGSGCLPVQSVFFRCVCPGNGGFCRPTPSAAPEVP 1384
Qy 890 GTTATTARACAA 901
Db 1385 EEPRCPRACQA 1396

RESULT 14
NTCL RAT
ID NTCL1 RAT STANDARD; PRT; 2531 AA.
AC C07006;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
```

```
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
RL development.";
RN Development 113:199-205(1991).
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP FUNCTION.
RX MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RN Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RL neural progenitor cells to an astroglial fate.";
RN Neuron 29:45-55(2001).
RP [4]
RT TISSUE SPECIFICITY.
RX MEDLINE=92302015; PubMed=1295745;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene.";
RN Development 116:931-941(1992).
RP [5]
RT TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RL functional roles for the Notch-DSL signaling system during brain
RN development.";
RX J. Comp. Neurol. 436:157-181(2001).
RP [6]
RT FUNCTION: Functions as a receptor for membrane-bound ligands
RX Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC [7]
RT SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
RX terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC [8]
RT SUBCELLULAR LOCATION: Type I membrane protein. Following
RX proteolytical processing NICD is translocated to the nucleus (By
CC similarity).
CC [9]
RT TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
RX Expressed in postnatal central nervous system (CNS) germinal zones
CC and, in early postnatal life, within numerous cells throughout the
CC CNS. Found in both subventricular and ventricular germinal zones.
CC [10]
RT DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
RX days 12 and 14 and decrease rapidly to much lower levels in the
CC adult.
CC [11]
RT PTM: Synthesized in the endoplasmic reticulum as an inactive form
RX which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC [12]
RT PTM: Phosphorylated (By similarity).
RX [13]
RT SIMILARITY: Belongs to the NOTCH family.
CC [14]
RT SIMILARITY: Contains 36 EGF-like domains.
CC [15]
RT SIMILARITY: Contains 3 Lin/Notch repeats.
CC [16]
RT SIMILARITY: Contains 5 ANK repeats.
CC [17]
RT This SWISS-PROT entry is copyright. It is produced through a collaboration
RX between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```


PROSITE; PS00022; EGF 1; 33.	FT	DISULFID	208	224	BY SIMILARITY.
PROSITE; PS01186; EGF 2; 26.	FT	DISULFID	226	235	BY SIMILARITY.
PROSITE; PS00026; EGF 3; 34.	FT	DISULFID	242	253	BY SIMILARITY.
PROSITE; PS01187; EGF 4; 16.	FT	DISULFID	247	262	BY SIMILARITY.
Receptor; Transcription regulation; Activator; Differentiation;	FT	DISULFID	264	273	BY SIMILARITY.
Developmental protein; Repeat; ANK repeat; EGF-like domain;	FT	DISULFID	280	293	BY SIMILARITY.
Transmembrane; Glycoprotein; Signal; Phosphorylation.	FT	DISULFID	287	302	BY SIMILARITY.
SIGNAL 1 40	FT	DISULFID	304	313	BY SIMILARITY.
CHAIN 41 2319	FT	DISULFID	320	331	BY SIMILARITY.
CHAIN 1631 2319	FT	DISULFID	325	340	BY SIMILARITY.
	FT	DISULFID	342	351	BY SIMILARITY.
CHAIN 1664 2319	FT	DISULFID	357	368	BY SIMILARITY.
	FT	DISULFID	362	379	BY SIMILARITY.
	FT	DISULFID	381	390	BY SIMILARITY.
	FT	DISULFID	397	410	BY SIMILARITY.
	FT	DISULFID	404	419	BY SIMILARITY.
	FT	DISULFID	421	430	BY SIMILARITY.
	FT	DISULFID	437	448	BY SIMILARITY.
	FT	DISULFID	442	457	BY SIMILARITY.
	FT	DISULFID	459	468	BY SIMILARITY.
	FT	DISULFID	475	486	BY SIMILARITY.
	FT	DISULFID	480	495	BY SIMILARITY.
	FT	DISULFID	497	506	BY SIMILARITY.
	FT	DISULFID	513	524	BY SIMILARITY.
	FT	DISULFID	518	533	BY SIMILARITY.
	FT	DISULFID	535	544	BY SIMILARITY.
	FT	DISULFID	551	561	BY SIMILARITY.
Query Match 9.8%; Score 562.5; DB 1; Length 2319;					
Best Local Similarity 22.7%; Pred. No. 1.7e-21;					
Matches 316; Conservative 33; Mismatches 482; Indels 561; Gaps 82;					
QY	38 GAAAGTC---GCCAGTGC	QY	38 GAAAGTC---GCCAGTGC	QY	38 GAAAGTC---GCCAGTGC
DB	38 GAAAPPCLDGSFCANGGRCTHQPSREAACLCLPLGTVGERCOLDPCHSPCARGVQCS	DB	38 GAAAPPCLDGSFCANGGRCTHQPSREAACLCLPLGTVGERCOLDPCHSPCARGVQCS	DB	38 GAAAPPCLDGSFCANGGRCTHQPSREAACLCLPLGTVGERCOLDPCHSPCARGVQCS
QY	71 CGCGG---GCCAGTGC	QY	71 CGCGG---GCCAGTGC	QY	71 CGCGG---GCCAGTGC
DB	98 SVVAGVAFRCRLGRGPDCLPFCFSPCAHGAAPCVSGSDRYACA--CPFGYQGRN	DB	98 SVVAGVAFRCRLGRGPDCLPFCFSPCAHGAAPCVSGSDRYACA--CPFGYQGRN	DB	98 SVVAGVAFRCRLGRGPDCLPFCFSPCAHGAAPCVSGSDRYACA--CPFGYQGRN
QY	118 C-----GGCGAGGT-----GGTGGCTGTCTG-TCTG--AGCAGACGTGGCC--	QY	118 C-----GGCGAGGT-----GGTGGCTGTCTG-TCTG--AGCAGACGTGGCC--	QY	118 C-----GGCGAGGT-----GGTGGCTGTCTG-TCTG--AGCAGACGTGGCC--
DB	157 CRSIDECRAGASCRHGGTCINTPGSFHCLPLGLCNPIVPCAPSCRNGGTQRQ	DB	157 CRSIDECRAGASCRHGGTCINTPGSFHCLPLGLCNPIVPCAPSCRNGGTQRQ	DB	157 CRSIDECRAGASCRHGGTCINTPGSFHCLPLGLCNPIVPCAPSCRNGGTQRQ
QY	159 -ATCTCGGCTGCGCGGCGC-----GCCGGGGCGCGCTGCC-----TGCCCT--	QY	159 -ATCTCGGCTGCGCGGCGC-----GCCGGGGCGCGCTGCC-----TGCCCT--	QY	159 -ATCTCGGCTGCGCGGCGC-----GCCGGGGCGCGCTGCC-----TGCCCT--
DB	217 SSDTYDCACLPFGEGQNCENVDDCFCHRCCLNGTCTVDGVTYNTQCPEPWTQFCTED	DB	217 SSDTYDCACLPFGEGQNCENVDDCFCHRCCLNGTCTVDGVTYNTQCPEPWTQFCTED	DB	217 SSDTYDCACLPFGEGQNCENVDDCFCHRCCLNGTCTVDGVTYNTQCPEPWTQFCTED
QY	200 ---GCTGGAC--GAGCAGCAGGTAAGCTGCTGCTCT-----ACGACATGAACGGGTG	QY	200 ---GCTGGAC--GAGCAGCAGGTAAGCTGCTGCTCT-----ACGACATGAACGGGTG	QY	200 ---GCTGGAC--GAGCAGCAGGTAAGCTGCTGCTCT-----ACGACATGAACGGGTG
DB	277 VDECQLQPNACHNGTGFNLLGHSVCVNGWTGSCSQNIDDCATAVCFHGATCHDRVA	DB	277 VDECQLQPNACHNGTGFNLLGHSVCVNGWTGSCSQNIDDCATAVCFHGATCHDRVA	DB	277 VDECQLQPNACHNGTGFNLLGHSVCVNGWTGSCSQNIDDCATAVCFHGATCHDRVA
QY	248 TTACTCAGC-----CCTCAGGAGCTGTGTGCC	QY	248 TTACTCAGC-----CCTCAGGAGCTGTGTGCC	QY	248 TTACTCAGC-----CCTCAGGAGCTGTGTGCC
DB	337 SFYCACPMGKTGLLCHLDACVSNPCHDAICDTPNVSRAICTCP---PGTGG-----	DB	337 SFYCACPMGKTGLLCHLDACVSNPCHDAICDTPNVSRAICTCP---PGTGG-----	DB	337 SFYCACPMGKTGLLCHLDACVSNPCHDAICDTPNVSRAICTCP---PGTGG-----
QY	276 AC-----CCTG-----CCCAGAACCGCAGGTGAGC-----AAGG	QY	276 AC-----CCTG-----CCCAGAACCGCAGGTGAGC-----AAGG	QY	276 AC-----CCTG-----CCCAGAACCGCAGGTGAGC-----AAGG
DB	389 ACDQVDSCISGANPCHEHLGRCVNTQGSFLQCGRGYTPRCETDVNCLSGPCRNQATC	DB	389 ACDQVDSCISGANPCHEHLGRCVNTQGSFLQCGRGYTPRCETDVNCLSGPCRNQATC	DB	389 ACDQVDSCISGANPCHEHLGRCVNTQGSFLQCGRGYTPRCETDVNCLSGPCRNQATC
QY	307 TGGAGATTCTCAG-----CACGTCATCGACT-----ACATCAGG--GACC--	QY	307 TGGAGATTCTCAG-----CACGTCATCGACT-----ACATCAGG--GACC--	QY	307 TGGAGATTCTCAG-----CACGTCATCGACT-----ACATCAGG--GACC--
DB	449 LDRIGQFTICINAGFTGTFCEVDIDECQSSPCVNGVCKDRVNGFSC--TCPSGFSGSTCQ	DB	449 LDRIGQFTICINAGFTGTFCEVDIDECQSSPCVNGVCKDRVNGFSC--TCPSGFSGSTCQ	DB	449 LDRIGQFTICINAGFTGTFCEVDIDECQSSPCVNGVCKDRVNGFSC--TCPSGFSGSTCQ
QY	346 ---TTCACT--TGGAGCT---GAACTC---GGAATCCG-----AAGTTGGAC-----	QY	346 ---TTCACT--TGGAGCT---GAACTC---GGAATCCG-----AAGTTGGAC-----	QY	346 ---TTCACT--TGGAGCT---GAACTC---GGAATCCG-----AAGTTGGAC-----
DB	508 LVDDECASTPCRNAGAKVQDPDGYECRCAGEGFTLCERNVDDCSFDPCHHGRVCVDGIAS	DB	508 LVDDECASTPCRNAGAKVQDPDGYECRCAGEGFTLCERNVDDCSFDPCHHGRVCVDGIAS	DB	508 LVDDECASTPCRNAGAKVQDPDGYECRCAGEGFTLCERNVDDCSFDPCHHGRVCVDGIAS
QY	383 ---CCC-----CGGGGGC-----CGAGGGCTG-----CC	QY	383 ---CCC-----CGGGGGC-----CGAGGGCTG-----CC	QY	383 ---CCC-----CGGGGGC-----CGAGGGCTG-----CC
DB	568 FSCACAPGYTGIRCESQVDECRSPCRVGGKCLDLDVKYLCRPGTGTGNCVNVDDCA	DB	568 FSCACAPGYTGIRCESQVDECRSPCRVGGKCLDLDVKYLCRPGTGTGNCVNVDDCA	DB	568 FSCACAPGYTGIRCESQVDECRSPCRVGGKCLDLDVKYLCRPGTGTGNCVNVDDCA
QY	404 GGTC---CGGGCTCGCTC-----AGCACCTCAACGG--CGAGATCAG-----CGC--	QY	404 GGTC---CGGGCTCGCTC-----AGCACCTCAACGG--CGAGATCAG-----CGC--	QY	404 GGTC---CGGGCTCGCTC-----AGCACCTCAACGG--CGAGATCAG-----CGC--

```

Db 796 CSCPQWQPRCQDVECA-----GPACGPHGICITNLAGSFCTCHGYTGPSCDDIDN 851
Qy 511 -C-----CAGGACCGCGGACCCAGCCATCCAGGGGCAAGAGGAATACGTCCTCG- 564
Db 852 DCDNPNCLNGSGCQDGVGFSCLPGFA-----GPRCARDVDECLSNPCGPG-TCTDH 904
Qy 565 TGGGTCTCCCCCAACGC-----GCCTGCGCGGATCTGAGGAGAACAA-AGACCGATCG 616
Db 905 VASFTCTCPGPGFCEQDLDCPSGSCFNGTCTVDGVNSFSLCRPGYTGACQHEAD 964
Qy 617 GC-----GGCCACTGCGCCCT-TAATCG-----CATCCAGC-----CTGGGGTGTAG- 657
Db 965 PCLSRPCLGGVCSAAHPFRCTCLESTGTPGQCQLVDWCSRQPCNGRCVQTGAYCLC 1024
Qy 658 -GCTG-----AGG-----CACTGG----- 670
Db 1025 PPQWSRELDIRSLPCREAAAGIVRLBQLCAGGCGQVDESSHYVCVPERGTSGHCEQE 1084
Qy 671 -----CGAGAGAG--GGCGCTC-----CTCTGCGACACC----- 699
Db 1085 VDFCLAPQCQHGTCRGYMGYMCCLPGYNDNCDDVDVDCASQPCQHGSGCIDLVARY 1144
Qy 700 -TACTAGTCA-----C-----CAGAGACTTTAGGGGTGGGATTCACATC 738
Db 1145 LSCSPGTLGVLCINEDDCGPPLDSGPRLNGTCTVDLVGG-----FRCTC 1193
Qy 739 GTGTGTTTCTATTTTGAAGAC-AGACATTTTAAAAATGG--TCAGTTTGGTGCTT 795
Db 1194 PPGYTLRCEADINE-----CRSGACHAAHTRDCLQDPGGGFRCLCHAGFSGPRCQT 1245
Qy 796 -----CTCAGATTTCTGAGGAATGCTTTGTATTGATTATTAATGATCACCGA 846
Db 1246 VLSPCESQFQHGQCRRSPFGG-----GLT-----FCHCAQP 1280
Qy 847 CTGAGAAATTTGTTTACAAATGATGCTGTGGG-----GCTGTTTTTGT-TATPAAA 898
Db 1281 FNGPRCVRARSRELQC--PVGVPCCQTPRGPACPPHSGPSRFSFGSPGASNAS 1338
Qy 899 CAAATAA-----TTTAGATGCTGAAGAAAAA 925
Db 1339 CAAAPCLHGSCRPAPLAPFRACAAQWTPRCEAPAAA 1378

RESULT 13
NTC3_RAT
AC NTC3_RAT STANDARD; PRT; 2319 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NTC3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Rattus norvegicus (Rat).
RT "Rattus norvegicus mRNA for Notch 3."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP FUNCTION.
RX MEDLINE=21094508; PubMed=1182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
RA Honjo T.;
RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
RT neural progenitor cells to an astroglial fate."
RL Neuron 29:45-55(2001).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;

```

```

RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
RT functional roles for the Notch-DSL signaling system during brain
RT development."
RL J. Comp. Neurol. 436:167-181(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(E)C which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
CC (CNS) germinal zones and, in early postnatal life, within
CC numerous cells throughout the CNS. It is more highly localized
CC to ventricular germinal zones.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(E)C. Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 34 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF164486; AAD4653.2; -.
CC HSP; P00740; IEDM.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR000152; Asx_hydroxyl_S.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR001438; EGF_II.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR008297; Notch.
CC InterPro; IPR000800; Notch_dom.
CC Pfam; PF00023; ank; 6.
CC Pfam; PF00066; notch; 3.
CC Pfam; PF00008; EGF; 33.
CC PRINTS; PIRSF002279; Notch; 1.
CC PRINTS; PR00010; EGFLOOD.
CC PRINTS; PR00011; EGFELAMININ.
CC PRINTS; PR01452; NOTCH.
CC SMART; SM00248; ANK; 6.
CC SMART; SM00179; EGF_CA; 20.
CC SMART; SM00004; NL; 3.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS50088; ANK_REPEAT; 4.
CC PROSITE; PS00010; ASX_HYDROXYL; 18.

```

or send an email to license@sib.ch).

```

CC  -----
CC  EMBL; U97669; AAB91371.1; --
DR  EMBL; AF058900; AAC14346.1; --
DR  EMBL; AF058881; AAC14346.1; JOINED.
DR  EMBL; AF058882; AAC14346.1; JOINED.
DR  EMBL; AF058883; AAC14346.1; JOINED.
DR  EMBL; AF058884; AAC14346.1; JOINED.
DR  EMBL; AF058885; AAC14346.1; JOINED.
DR  EMBL; AF058886; AAC14346.1; JOINED.
DR  EMBL; AF058887; AAC14346.1; JOINED.
DR  EMBL; AF058888; AAC14346.1; JOINED.
DR  EMBL; AF058889; AAC14346.1; JOINED.
DR  EMBL; AF058890; AAC14346.1; JOINED.
DR  EMBL; AF058891; AAC14346.1; JOINED.
DR  EMBL; AF058892; AAC14346.1; JOINED.
DR  EMBL; AF058893; AAC14346.1; JOINED.
DR  EMBL; AF058894; AAC14346.1; JOINED.
DR  EMBL; AF058895; AAC14346.1; JOINED.
DR  EMBL; AF058896; AAC14346.1; JOINED.
DR  EMBL; AF058897; AAC14346.1; JOINED.
DR  EMBL; AF058898; AAC14346.1; JOINED.
DR  EMBL; AF058899; AAC14346.1; JOINED.
DR  EMBL; AF058900; AAC14346.1; JOINED.
DR  EMBL; AC004257; AAC04897.1; --
DR  EMBL; AC004563; AAC15789.1; ALT_INIT.
DR  PIR; S78549; S78549.
DR  HSP; P00740; IEDM.
DR  Genew; HGNC:7883; NOTCH3.
DR  MIM; 600276; --
DR  MIM; 125310; --
DR  InterPro; IPR002110; ANK.
DR  InterPro; IPR000152; Asx_hydroxyl_s.
DR  InterPro; IPR000742; EGF-2.
DR  InterPro; IPR001881; EGF-Ca.
DR  InterPro; IPR001438; EGF-II.
DR  InterPro; IPR006209; EGF-like.
DR  InterPro; IPR002049; Laminin_Egf.
DR  InterPro; IPR008297; Notch.
DR  InterPro; IPR000800; Notch_dom.
DR  Pfam; PF00023; ank; 6.
DR  Pfam; PF00008; EGF; 34.
DR  Pfam; PF00066; notch; 3.
DR  PRINTS; PRS002279; Notch; 1.
DR  PRINTS; PRS00010; EGF-BLOOD.
DR  PRINTS; PRS00011; EGF-LAMININ.
DR  PRINTS; PRS00012; NOTCH.
DR  SMART; SM00248; ANK; 6.
DR  SMART; SM00179; EGF_CA; 19.
DR  SMART; SM00004; NL; 3.
DR  PROSITE; PS00297; ANK_REPEAT; 1.
DR  PROSITE; PS00088; ANK_REPEAT; 4.
DR  PROSITE; PS00010; ASX_HYDROXYL; 18.
DR  PROSITE; PS00022; EGF_1; 33.
DR  PROSITE; PS01186; EGF_2; 25.
DR  PROSITE; PS00026; EGF_3; 34.
DR  PROSITE; PS01187; EGF_CA; 16.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW Disease mutation.
FT SIGNAL 1 39 POTENTIAL.
FT CHAIN 40 2321 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT CHAIN 1629 2321 NOTCH EXTRACELLULAR TRUNCATION (BY
FT CHAIN 1662 2321 SIMILARITY).
FT CHAIN 1662 2321 NOTCH INTRACELLULAR DOMAIN (BY
FT CHAIN 40 1643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1644 1664 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1665 2321 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 40 77 EGF-LIKE 1.
FT DOMAIN 78 118 EGF-LIKE 2.
FT DOMAIN 119 156 EGF-LIKE 3.
FT DOMAIN 158 195 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).

```

```

FT DOMAIN 197 234 EGF-LIKE 5.
FT DOMAIN 236 272 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 274 312 EGF-LIKE 7.
FT DOMAIN 314 350 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 351 389 EGF-LIKE 9.
FT DOMAIN 391 429 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 431 467 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 469 505 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 507 543 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 545 580 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 582 618 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 620 655 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 657 693 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 695 730 EGF-LIKE 18.
FT DOMAIN 734 770 EGF-LIKE 19.
FT DOMAIN 771 808 EGF-LIKE 20.
FT DOMAIN 810 847 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 849 885 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 887 922 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).

Query Match 9.9%; Score 570.5; DB 1; Length 2321;
Best Local Similarity 23.6%; Pred. No. 6.7e-23;
Matches 321; Conservative 26; Mismatches 496; Indels 517; Gaps 79;

QY 20 CCAGTCGCCAAGATCATGAAAGT-----CGCCAGTGG-----CAGCACCG-- 60
DB 82 CHSGPC-----AGRGVCCSSVVAGTARFSCPRGFRGDCSLDPCLSSPCAHCARCSVG 137
QY 61 -----CCAC-----CGCCGCGCGGGCCCGAGTGGC-----GCTGAA----- 94
DB 138 PDGRFLSCPPGYGGRSRSRDVDECRVGEPCRHGGTCLNTGFSFRCPCAGYTGFLCENP 197
QY 95 -----GGCC--GGCAAGACAGCAGCGGTG-----GGCGGAGGT----- 127
DB 198 AVPCAPSPCRNGGTCROSGDLTYDCA-CLPFGFQONCEVNVDDCPGHRCLNGGTCVDG 256
QY 128 -----GGTGGCCT-----GTCCTCTGAGCAGAGCGTG--GC----- 157
DB 257 TYNCCPPPEWTGQFCTEDVDEBCLQPNACHNGGTGFTNLGHSVCVNVGMTGESCSQ 316
QY 158 -CATCTC--GGCTGCGCGGGCGC--CGGGGCGCGCTGCTGCG----- 197
DB 317 DCATVCFHGATCHDRVASFYCACPMGKTGLLCHLDACVSNPCHEPAICDTNPVNGRAI 376
QY 198 CT---GCTGGAC-----GAG-----CAGCAGGTAAACGTGTGCTCTA----- 232
DB 377 CTCPPGTGGACDQDVDECSIGANPCBHLGRCVNTQGSFLCQCGRYTGPRCETDVNECL 436
QY 233 CGACATGAACGGCTGTTACTC-----ACGCTCAAGGAGC--TGCTGCC 275
DB 437 SGPCRNQATCLDRIGQFTCIOMAGFTYCEVIDECQSSFCVNGGV-CKDRVNGFSCTC 495
QY 276 ACCCTGCC-----CA-----GAAC-----CGCAAGGTGAGCAAG----- 305
DB 496 PSGFSGSTCQLDVEDECASITPCRNAGAKVDQPDGVEGRCABGFEGLCDRNVDDCGSPDCH 555
QY 306 ---GTGAGATTCTCCAGCAGGTACATCATATCAGGAC-----CTTCAGTTG-- 353
DB 556 HGRCVNDGIASFSCACAPGYTGTRCESQVDECRSOPCRHGGKCLDLVDKYLKCRCPSGTGV 615
QY 354 -----GAGCTGAATCG-CAATCCGAAGTTGG-----GACCCC--CGGGGGCGCGA 395
DB 616 NCEVNIDDCASNPTCTFGVCRDGINRYDVCQPGTGPLCNVINECASSPCGEGSCVDG 675
QY 396 GGG-----CT--GCGGTC--CGGGTC--CGCTCAGCACCTCA-- 429
DB 676 ENGRFCLCPPGSLPPLCLPSPHPCAHFPCSHGICYDAPGFRFCVCEPFGWGSQSLAR 735
QY 430 -ACGG--CGAGATCAGCG-----CCTGACGG--C-----CGAGGC--GGCATCGTTCC 472
DB 736 DACESQPCRAGTCTSSDGMGFHCTCTCPFGVQRCQCELLSPCTNPCEHGRCESAPGQLPV 795
QY 473 TGC--GGACGATC-----GCATCTTGTTGTCG-----CTGAAAGGCGCTC-----CC----- 510

```


SQ	SEQUENCE	801 AA; 65407 MW; EA54C9BF4A5A0F41 CRC64;
	Query Match	10.9%; Score 627; DB 1; Length 801;
	Best Local Similarity	28.5%; Pred.No.4.7e-25;
	Matches	Conservative 24; Mismatches 337; Indels 342; Gaps 44;
Qy	2	GGGCCCATCTCTGTTTC-----AGCCAGT---CGCCAAGAATCATGAAGTCGCACGTGG 52
Db	83	GGGAYAAEAANAASPLLIAPINAQFVAATGRFLIENGANGAP--GTGANGGPGWLIINGG 140
Qy	53	CAGCACCGCCACCGCGCGCGGGCCCCAGCTGCGCGCTGAAGCGCGCAAGACAGC---109
Db	141	AGGSAPG---AGAGNGGAGG---LFGSGAGGASTDVAGGAGGAGGAGNAGMLFG 192
Qy	110	GAGCGGTG-----CGGGCAGGTGG-----TGCCCTCTCTGTCTGACACAGCGTG 155
Db	193	AAGVGVGVGFNGGATGAGGAGGAGGLFGAGRERSGSGNLTG---GAGGAG---GNA 246
Qy	156	GCCATCTCGCCT--GCCGGGGCGCGGGCGCGCTGCTGCCCTGCTGGA-----CGA 208
Db	247	GTLATDGGAGTGGASRSGFGGAGGAGGADAGMPFG-----SGSGGAGGISKSVGD 299
Qy	209	GCAGCAGATAACGTGCTCTACACATGAACGGCTGTACTCAOCCTCAAAGAGC- 267
Db	300	SAAGGAGGAPGLIINGNG-----GNNGASTGGDGDPG-----AGGTGVL 341
Qy	268	--TGGTGCCACCCTGCCCCAGAACCGCAAGCTGAGCAAGTGGAGATTCTCCAGCAGT 325
Db	342	IGNNG-----GSGGTGATLGKALGGT-----GV 368
Qy	326	CATCGACTACATCA-----GGGACCTTCAGTTGAGCATG 359
Db	369	LLGLDFTAPASTSPHLTLQQDVINVMNDPFTLTGRFLIENGANGTPGTGADGGAG--G 426
Qy	360	AACTCGGAATCGAAGTTGGNACCCCGGGGCGCAGGGCTGCCGCTCGGGCTCCGTC 419
Db	427	WLFNGEN-----GGGGTITGG-----VNGGAGGAGGAGGILFG-TGGT---GSGGGPGATG 473
Qy	420	AGCACCTTCAACGGCGAATCAGCCCTTGACGGCGGAGGGCGCATGCTTCTCGCGAC 479
Db	474	LG-----GUGAGGAALLFG-----SGGAGSGGAGAVGNG-----GAGNA 511
Qy	480	GATCGCATCTGTGTCGTGAAGCGCCTCCCACGGACCGCGCGACCCCGACCATCAG 539
Db	512	GALLGAA-----GAGGAGGAG-----AVGN---GGAGN---GGLFANGAG 548
Qy	540	GGGGCAAGAGGAATTAAGTGTCTCTGTGGTCTCCCCCAACGCGCTCGCCGATCTGAG 599
Db	549	FPGGPSFAGGICGAGGNGGLFGAG-----TGAGGGSTLAGAG 591
Qy	600	GAGAACAGACGATCGCGGCCCATCGCGCCTTAACATGCAATCCAGCTGGGGCTGAGGC 659
Db	592	GAG---GNGGLPGA--GGTG---GAGSHSTAAGVSG-----GAGGA 624
Qy	660	TGAGGCACTGGCGAGGAGGCGCTCTCTCTGCACACCTACTAGTACCAGAGACTTT 719
Db	625	GDAGLLSLG--ASGAGGSGSSLT-----AAGVVGGTGGAGGLLFG 665
Qy	720	AGGGGTGGGATTCACACTCGTGTGTTCTATTTTTTGAAAAGCAGACATTTTAAAAAATG 779
Db	666	SGAGCGG-----FSNSGNGGAGGAGGADGLLVGSGAGGA 702
Qy	780	GTACGTTTGGTCTCTCAGATTCTGAGGAATTGCTTGTATTGTATTATTAACAATGA 839
Db	703	GASATGAATGGD-----GAGGKSAGP----- 725
Qy	840	TCACCGCATGAAATATGTTTTTACAATAGTTCT--GTGGGGCTGTTTTTTTATTATAA 897
Db	726	----LGDDGAGGATGLSGAFHIGKGVGGSVLIINGENGNG-----GNSGN---- 771
Qy	898	ACAAATAATTTAGATGGTGA AAA 920
Db	772	-----AGKSGAPGPSGAGA 787

```

RESULT 11
PHX5_MOUSE STANDARD; PRT; 672 AA.
ID PHX5_MOUSE
AC P08399;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Per-hexamer repeat protein 5.
GN PHX5 OR PER.
OS MUS MUSCULUS (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014384; PubMed=24113365;
RA Shin H.S., Bargiello T.A., Clark B.T., Jackson F.R., Young M.W.;
RT "An unusual coding sequence from a Drosophila clock gene is conserved
in vertebrates.";
RL Nature 317:445-448 (1985).
CC -1- SIMILARITY: TO OTHER G-T STRETCH CONTAINING PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M12039; AAA88320.1; -.
CC EMBL; X02966; CAA26710.2; -.
CC PIR; A24403; UNMK5.
CC MGD; MGI:104521; Phxr5.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR006210; IEGF.
CC SMART; SM00191; EGF; 1.
CC DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
CC Repeat.
CC DOMAIN
CC SEQUENCE 672 AA; 57924 MW; B85BF428CF424C0B CRC64;
CC
Query Match 10.7%; Score 618.5; DB 1; Length 672;
Best Local Similarity 31.4%; Pred. No. 1.le-24;
Matches 266; Conservative 15; Mismatches 321; Indels 243; Gaps 39;
QY 117 GCGGGCCAGGTGGTGGCG-----TGT 137
DB 6 GCGHICKEAPRGSVACRPFELAKNQKDCIKVELSISEVGTAYTEVRKTTGTATGI 65
QY 138 CTGTCTGAG---CAGACGTGGCCATCTCGGCTCGCGGGCGCGGGCGCGCTCGCT 194
DB 66 ATGTCCTGTRVKATGRGTGTDPTGTVTASATVATVARTVGTGTGTATVTEGTAKVTDGT 125
QY 195 GCCCTGCTGCAGCAGCAGCAG-GTAAACGTGCTGCTCTACGACATGAACGGCTGTACTC 253
DB 126 G-----TGTAKVTGTAKVGTGTGTGTGTG-TG-----TGTGTGTGTAKVT- 167
QY 254 ACGCCTCAAGAGCTGGTGTGCCACCCCTGCCCCAGAACCGCAAGGTGAGCAAG-GTGGAGA 312
DB 168 -----GTGTDRTG-----TGT-----GTGTGTGTGTGTGTAKV 196
QY 313 TTCTCCAGCAGTCATCGATACATCAGGACCTTCAGTTGGAGCTGAACTCGGAATCCG 372
DB 197 T-----GTAKVTGTGTAKVGTG-----TGTGTGTGTG-TGT-----GTDGTG 236
QY 373 AAGTTGGACCCCGGGGGCGGAGGCTCGGCTCGGGCTCCGCTCAGCACCTCAACG 432
DB 237 TAKVTGTGT-----GTGTG-TGTGTGTGTGTG-----TGTAKVTGTGTDR 275
QY 433 GCGAGATCAGCGCCTCAGCCGCGAGCGGCATGCGTCTCTCGGGAGCATCGCATCT-TG 491

```

```
DR Tuberculin; RV2634C; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SQ SEQUENCE 778 AA; 63131 MW; DAB20F58E4999E7 CRC64;

Query Match 10.9%; Score 629; DB 1; Length 778;
Best Local Similarity 31.4%; Pred. No. 3.7e-25;
Matches 268; Conservative 8; Mismatches 325; Indels 252; Gaps 37;

QY 37 TGAAGTCCCATGTCGACACCGCCACCGCGCGCGCGCCAGCTGCGCGCTGAAG- 95
DB 128 TGAPGPGGGLLNGNGG-----SGAPQPGAGDAGLNG-CTGKGGD 174
QY 96 GCGCGCAAGACACGCA---GCGGT-CGCGGCGA---GGTGTGCGCTGTC-TGCTGAGC 147
DB 175 GLVSGAAGGAGGVRGGLLNGGTGGAGGAGATLVGGTGGVG-GATGLIGSGFGGAGG 233
QY 148 AGACGTGGCCATCTCGCGTGCCTGCGGGGCGCGCGCGCGCGCTGCTGCTGCGAC 206
DB 234 AAAGVGTG-----GVGSGGVGVGVENG-----FGAG 263
QY 207 GAGCAGCAGGTAAACGTGCTCTACGACATGACCGCTGTATTCTACCGCTCAAGGAG 266
DB 264 GLAAGGVGAASVFTGGG-----GVGDGAP-GDGGAGPLLNG-GVGLGGAG 315
QY 267 CTGTGCGCCACC---TGCCCCAGAACCGCAAGGTGAG-----CAAGTGAGATTCT 316
DB 316 AAGNGGAGGMLLDGGAGGCGGPAVAGVLGGMPGAGNGNANWFGGAGGQGGT--- 372
QY 317 CCAGCAGCTCATCGACTACATCAGGACCTTCACT---TGGAGCTGAATCGGAATCCGA 373
DB 373 -----GLAGTNGVNPGSIANPNTGANGTDSNGNGTGGNGPGEA---GGVAGGV 422
QY 374 AGTTGGACCCCGGGGCGGAGS---CTCTCCCGTCCGGCTCGCTCAGCCCTCAC 431
DB 423 GGQGLGESLDNGTGTGKGAGGTAGTDGGAG-----AG 458
QY 432 GCGCAGATCAGCGCCCTCAGCGCGGCGGCGGCTGCTCTCCGCGACGATCGCATCTTG 491
DB 459 GAGIGETDSAGVATGEGGAGATGVDGVGGAG---GKGGQ-GHNTGVGDAFG 512
QY 492 TGTCCGTGAAGCGCTCCCGCAGGACCGGCGGACCCCGACCCATCCAGGGGCAAGGA 551
DB 513 DGGTGGDNGALG-----AAGNGGTGGAG-----NGRGGMLIANGGA 552
QY 552 ATTAGCTGCTGTGGTCTCTCCCAACGCGCTCGCGGATCTGAGGGAGAACAGACC 611
DB 553 G-GAGGTG-----GTGG-----GAGFAGGVGGAGGELTDGAGTAE 589
QY 612 GATCGCGGCCACTCGCGCCCTTAATCATCATCCAGCTGGGCTGAGCTGAGCACT--- 668
DB 590 GGT-GGLG-----LGGVGTGGMGSGSGVGVNGGAGSLIG 625
QY 669 --GCGAGGAGGSGCGCTCTCTCTGCAACCTACTAGTACCAGAGATTAGGGGGT 726
DB 626 LGGGGGAGGVGGTGGIG-----GUGG-----AGNGGA 653
QY 727 GG-GATTCCATCGCTGTGTTTCTATTATTTTAAAGCAGACATTTTAAATAATGGTACG 785
DB 654 GGAGTTTGGGATIGGGGT-----GGVGGAGGTGGTGGAGTTGGSGGAG 698
QY 786 TTTGGTGTCTCAGATTCTGAGGAATTCCTTTGTTATTTATATATACATGATCACCG 845
DB 699 GLIHWAGAGGTGAGGTGGGGLGGGNGG--NGGTGATGGGQDFFALGG----- 749
QY 846 ACTGAGATATTG 858
DB 750 ---GAGGAGGSGFG 759
```

RESULT 10

```
Y747_MVCTU STANDARD; PRT; 801 AA.
ID Y747_MVCTU
AC OS810;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PE-PGRS family protein RV0747/MT0772.5 precursor.
GN RV0747 OR MT0772.5 OR MT041.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruster S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oehkosh;
RC MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021958; CAA17514.1; -.
CC EMBL; AE006968; AAK45011.1; -.
CC FIR; F70824; F70824.
CC TIGR; MT0772.5; -.
CC Tuberculin; RV0747; -.
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30
FT CHAIN 31 801
FT CONFLICT 188 188 G -> S (IN REF. 2).
FT CONFLICT 225 225 R -> G (IN REF. 2).
FT CONFLICT 227 227 R -> G (IN REF. 2).
FT CONFLICT 295 295 K -> R (IN REF. 2).
FT CONFLICT 300 300 S -> G (IN REF. 2).
FT CONFLICT 338 338 T -> I (IN REF. 2).
FT CONFLICT 377 377 A -> P (IN REF. 2).
FT CONFLICT 577 577 T -> A (IN REF. 2).
FT CONFLICT 580 580 MISSING (IN REF. 2).
```

```
FT REPEAT 583 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT CONFLICT 662 665 V -> L (IN REF. 1).
FT CONFLICT 662 662 S -> T (IN REF. 1).
FT CONFLICT 672 672 NPGLSGCDVLIQALLLEWVALIOLGSSSIGOVNYSAGQA
FT CONFLICT 695 747 TQIVGQSVQAL -> ILVFLDWSSFKLFSRLFLLSRS
      (IN REF. 1).
FT SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;

Query Match 11.0%; Score 635.5; DB 1; Length 747;
Best Local Similarity 30.6%; Pred. No. 1.7e-25;
Matches 280; Conservative 21; Mismatches 358; Indels 255; Gaps 39;

Qy 30 AGAATCATGAA-----AGTCCGAGTGGCAGCACCACCGCGCGCGGCCCCAGCT 84
      |||
Db 3 AGAAAAAGAGGQGGYGLGQGGAGQGGYGLGQGG-----AGQAGAAAAAAG 52
      |||

Qy 85 GC-----GGCTGAAGCGCGGAAGACAGCAGCGAGCGGTGGG-----GCGAGGTGGTGCCTGT 137
      |||
Db 53 GAGGGYGLGSGQAGRGQGGAGAAAAAGAGGAGGQGGYGLGSGQAGRGGLGGQAG-----108
      |||

Qy 138 CTGTCTGAGCAGAGCTGGCCATCTCGCTGTCGCGGCGCGGCGCGGCGCGCTGCTGCC 197
      |||
Db 109 -----AAAAAAGGAGGQGGYGLGQGG-----RGQGAAGAAAGAGGQGGYGLGSGQGA 159
      |||

Qy 198 CTGTCTGAGCAGCAGCAGGTAAGCTGTCTCTACACATGAACGCGTGTACTCAQGC 257
      |||
Db 160 GRGSLGGQ-GGAAAAAAGAGGQGGY-----GLGQGGAGQGGYGLGSG-----202
      |||

Qy 258 CTCAGAGAGTGTGTCCCACTCTGCCCGAGAACCGCMAGTGTAGCAAGTGTGAGATCTC 317
      |||
Db 203 ---QGAGRGGLGGQ-----AGAA-AAAAAGGAG-----QGGILGGQ-----235
      |||

Qy 318 CAGCAGCTCATCGACTACAGGACCTTCAGTTGGAGCTGAACCTCGGAATCCGAAGTT 377
      |||
Db 236 -AGGAG-----ASAAAGAG-----QGGYGLGSGQAGRGGAGAAAAAGGA 280
      |||

Qy 378 GGGACCCCGGGCGGAGGCTCCCGTCCGGTCCGGCTCCGCTCAGCACCTCAACGGCGAG 437
      |||
Db 281 GGGY-----GGLGGQAGGQGGYGLGSGQAGRG-----GLGQGGAG 317
      |||

Qy 438 ATCAGGCCCTGAGCGCGGCGGCGATGGTCTCTCGGACGATCGATCTGTGTGCG 497
      |||
Db 318 AAAAG-----GA-----GQGLGGQAGGQ-----AGAAAAAGGA-----GQGGYGG 355
      |||

Qy 498 TGAAGCGCTCCCGGAGGACCGGCGGACCCAGCCATCCAGGCGGCAAGAGGAATTAAG 557
      |||
Db 356 LGSQAG-----RGLGGQ-----GAGAVAAAAAGGAG--QGGYGLGSGQAG 396
      |||

Qy 558 TGCTCTGTGGTCTCCCCCAAGCGCCCTCGCG--GATCTGAGG-GAGAACAGACCGAT 614
      |||
Db 397 RG-----QGAG-----AAAAAGGAGQGGYGLGQAGRGGLGGQAGAA-AAAAAGGAG 447
      |||

Qy 615 CGGCGGCCACTCGCGCCCTTAATCATCATCGCTGCGGCTGAGGCTGAGCACTGGCGAG 674
      |||
Db 448 QGGYGG-LGNQAG-----RGGQAAAGAGGAGQGG-----YGLGSGQ 484
      |||

Qy 675 GAGAGGGCGCTCTCTCTGACACTACTA-----GTCAACGAGACTTTAGGGGTGG 728
      |||
Db 485 GAGRGG-----QGAGAAAAAAGAGQGGYGLGSGQGGYGLGSGQGGRLGG 533
      |||

Qy 729 GATTCACCTCGTGTGTTCTATTTTGAAGACAGACATTTTAAAAAATGTCACGTTT 788
      |||
Db 534 QGAGAAAAAG-----GAGQGGILGGQAGGAGAAAAAGGAGVQGG--Y 575
      |||

Qy 789 GGTGCTCTCAGATTTCTGAGAAATGCTTTGTTGTTGTTATTTATTAATCATCACTG 848
      |||
Db 576 GGLG-----SQAGRGQGGAGAAAA-----AAGGAGQGGYGLG 609
      |||

Qy 849 GAGATATGTTTACATAGTCTGTGGGCTGTTTGTGTTTATTAACAAATATTT 908
      |||
Db 610 GQGVGRGLGGGAGAAAGAGGAGGQGGYGGV-SGASAAASAAARLSSPQASRVSSVSN 668
      |||
```

RESULT 9

```
YQ34_MYCTU
ID YQ34_MYCTU STANDARD; PRT; 778 AA.
AC F71933;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PE-PGRS family protein RV2634c/MT2712/MB2667c.
GN RV2634C OR MT2712 OR MTCY441.04C OR MB2667C.
OS Mycobacterium tuberculosis, and
OC Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Sulten S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
      complete genome sequence.";
      Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr. Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
      laboratory strains.";
      J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
      Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC CC
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
      SUBFAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (See http://www.isb-sib.ch/announce/
      or send an email to license@isb-sib.ch).
CC CC
CC EMBL; Z80225; CAB02341.1;
CC EMBL; AB007103; AAK47026.1; ALT_INIT.
CC EMBL; BX248343; CAD94852.1;
CC PIR; F70963; F70963.
CC TIGR; MT2712; --
```


	RV0278C/NT0291.
FT	M -> I (IN REF. 2).
CONFLICT	40
MISSING	163
R -> G (IN REF. 2)	807
SEQUENCE	957 AA; 81905 MW; 71EBABD41FBA47C CRO64;
Query Match	11.2%; Score 647.5; DB 1; Length 957;
Best Local Similarity	30.3%; Pred. No. 5.1e-26;
Matches 270;	Conservative 18; Mismatches 351; Indels 251; Gaps 34;
QY	37 TGAAGTCGCAGTGGGACGCCCGCCCGCGCCCAGCTGCGCGGTGAGG 96
DB	125 TGANGGDGMILNGSAGG-----SGAAGVNSG---AGGNG-GAGNGGAGG 167
QY	97 CGCGCAACACAG-CGACGGTGCGGG---CGAGGTGGTGGCTCTGTCTGTGACGAG 151
DB	168 LIENGCRGAGGVASSIGSSGGAGGNAMLFGAGGAGGAGGVV-ALTGGAGGAGGAGN 226
QY	152 GTGGCCATCTCGCGCTCCGG-----GGCGCCGGCGCGCTGCTCCCTGCTGTTG 204
DB	227 AG-----LLFGAAGVGGAAGFTNGSALUGAG--GAGGAGGLFATGTVGGSGGAGSGG 277
QY	205 ACGAGCACG---CAGGT-----AAACGTGCTGCTCTACGACATCAAACG-----GC 245
DB	278 AGGAGGAGLFGAGGTGGHGCFADSSFGVGGAGGAGGLFGAGGEGSGHSLVAGDG 337
QY	246 TGTTACTCACCCCTCAAGGACTGTGTCCACCCTCCCCAGAACGGCAAGTGAGACAAG 305
DB	338 AGSNAGMLALG----AAGGAGGTGGSD-----GTLFAGGTGGAGGAGGAGNALLFSG 385
QY	306 GTGGAGATTCTCCAGCACGTCAATCATACATCAGGACCTTCAGTTGAGCT-----G 359
DB	386 GSGGAGG--PGFADGGQGPPGNAGTVFSGSGAGNGGVGCGFAAGTAGGAGTPLLGN 443
QY	360 RACTCGGAATCCGAAGTTGGACCCCGCGGGCCGAGGGCTCCG-----404
DB	444 NGNGGGASAVTGGNGGTGGTVLI GNNGGSGSGGI GAKAGVGVSGLLLGDFNAPA 503
QY	405 -----GTCCGGGC-----TCGCG 417
DB	504 STSPHLTIQQNVLMWNEPOTLTGRPLI NGANGTPGTGADGAGGWLFNGANGTPGT 563
QY	418 TCAGCACCTCAACGGCGAGATCAGCGCCCTGACGGCCGA-----GGCGGCATGCGTT 470
DB	564 GAAGGAGGWLFPNGNGNHGHAT--NTAATATGGAGAAGGLFPTGGNGTGTGATGAG-- 619
QY	471 CTGCGGACGATCGCATCTGTGTGCTGTAAGCGCTCCCCAG--GGACCGCGGACCCC 529
DB	620 ---GIGGAGGA--GGVSLLISGGTGGNGNSIG-----VAGTGG--GGRGDAGLL 665
QY	530 ACCCATCCAGGGGCGACAGAGGAATTACGTCTGTGGGTCTCCCGACGGCGCTCGCC 589
DB	666 FGAAGTGGHGAAGVPFAGVGA-----GGN-----690
QY	590 GGATCTGAGGAGAACAGACCGATCGCGCGCCACTCGGCCCTTAAGTCATCCAGCGTG 649
DB	691 GGLFANGAGGAGGFNAAGG---NGNGSLFPTGTGTGAGTNFG-----AGNGG 737
QY	650 GGGCTGAGGCTGAGCACTGGCG--AGAGAGGGCGCTCCTCTCTGACACCTACTACTG 707
DB	738 NGLLPKAG---GTGGAGSGSGGILLTGGGHGNNAG-----LLSLGAS 777
QY	708 ACCAGAGACTTTAGGGGTGG-GATTCACATCGGTGTTTTCTATTATTTTAAAAACAGAC 766
DB	778 GGAGSGGASSLAGGAGGTGGNGALLFPRGAGGAGHGGAAUTSIQQGAGGAGGAGNL 837
QY	767 ATTTTAAAATAATGTCAGTTTGTGCTCTCAGATTCTTGAGGAATTG 816
DB	838 LFGSAGAGGAGGSANALGATGTTG-----GDGHHAGVF 873

RESULT 8
SPD1_NEPCL

RESULT 8
SPD1 NEPCL

ID	SPD1_NEPCL	STANDARD;	PRT;	747 AA.
PI9837;				
AC	01-FEB-1991	(Rel. 17, Created)		
DT	01-NOV-1995	(Rel. 32, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DT	Spidroin 1	(Dragline silk fibroin 1) (Fragment)		
DE	Nephila clavipes	(Orb spider)		
OS	Eukaryota: Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_TaxID=6915;			
RN	[1]			
RN	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RP	MEDLINE=90384959; PubMed=2402494;			
RP	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RL	[2]			
RN	SEQUENCE OF 653-747 FROM N.A.			
RP	MEDLINE=94165058; PubMed=8120021;			
RP	Beckwitt R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk			
RT	proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and			
RT	Araneus bicentarius (Araneidae).";			
RT	J. Biol. Chem. 269:6661-6663(1994).			
RL	[1]- FUNCTION: Spiders major ampullate silk possesses unique			
CC	characteristics of strength and elasticity. Fibroin consists of			
CC	pseudocrystalline regions of antiparallel beta-sheet interspersed			
CC	with elastic amorphous segments.			
CC	[1]- SUBUNIT: MAJOR SUBUNIT, WITH SPIROIN 2, OF THE DRAGLINE SILK.			
CC	[1]- SUBCELLULAR LOCATION: Extracellular.			
CC	[1]- DOMAIN: Highly repetitive protein characterized by regions of			
CC	polyalanine and glycine-rich repeating units.			
CC	[1]- SIMILARITY: belongs to the silk fibroin family.			
CC	[1]- DATABASE: NAME=Protein Spotlight;			
CC	NOTE=Issue 24 of July 2002;			
CC	WWW="http://www.expasy.org/spotlight/articles/splt024.html".			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL; M37137; AAA29380.1; -				
DR	EMBL; U03848; AAB60212.1; -			
DR	PIR; A36068; A36068.			
KW	Silk; Repeat.			
FT	NON TER	1		
FT	DOMAIN	1 655		
FT	REPEAT	1 25		
FT	REPEAT	26 38		
FT	REPEAT	39 66		
FT	REPEAT	67 96		
FT	REPEAT	97 130		
FT	REPEAT	131 158		
FT	REPEAT	159 191		
FT	REPEAT	192 204		
FT	REPEAT	205 235		
FT	REPEAT	236 262		
FT	REPEAT	263 292		
FT	REPEAT	293 305		
FT	REPEAT	306 333		
FT	REPEAT	334 360		
FT	REPEAT	361 394		
FT	REPEAT	395 424		
FT	REPEAT	425 458		
FT	REPEAT	459 485		
FT	REPEAT	486 512		
FT	REPEAT	513 525		
FT	REPEAT	526 555		
FT	REPEAT	556 582		
FT	REPEAT	583 610		
FT	REPEAT	611 640		
FT	REPEAT	641 670		
FT	REPEAT	671 700		
FT	REPEAT	701 730		
FT	REPEAT	731 760		
FT	REPEAT	761 790		
FT	REPEAT	791 820		
FT	REPEAT	821 850		
FT	REPEAT	851 880		
FT	REPEAT	881 910		
FT	REPEAT	911 940		
FT	REPEAT	941 970		
FT	REPEAT	971 1000		
FT	REPEAT	1001 1030		
FT	REPEAT	1031 1060		
FT	REPEAT	1061 1090		
FT	REPEAT	1091 1120		
FT	REPEAT	1121 1150		
FT	REPEAT	1151 1180		
FT	REPEAT	1181 1210		
FT	REPEAT	1211 1240		
FT	REPEAT	1241 1270		
FT	REPEAT	1271 1300		
FT	REPEAT	1301 1330		
FT	REPEAT	1331 1360		
FT	REPEAT	1361 1390		
FT	REPEAT	1391 1420		
FT	REPEAT	1421 1450		
FT	RE			

QY 151 GCGTGGCCATCTCGCGCTGCCGGGGC---GCCGGGGCGCGCTGCTGCGCC---TGCTG 203
 Db 493 TCENSLVLAICDXTSSRACIKWKKYKQCVLAS-----SATTTHADCTYHSTCTL 545
 QY 204 GACGAGCA---GCAGGT-AAAGCTGCTGCTCTACGACATGAACGGC---TGTTACTCAG 256
 Db 546 SNGSGTCVPLPKCEAITIENCLKANGOPCGWNGSQICDKACSTASKTFTTTGCTGCH 605
 QY 257 CTTCAAGGAGCTGGT---GC-----CCACCTGCGC-----CCAGAA 289
 Db 606 ISTCVANNPVTNGSLTIQGCPLTSCAARKSENCEIARVGFPTCLWVSSSTSCVEKS 665
 QY 290 CCGCAAGGTGAGCAAGGTGAGATTC---TCCA-----GC-----A 322
 Db 666 CATASTVGTGALSAGGFTFSGGQTYLNTICISNNTADGCIAPSSGSSILVSSNCRDGSKA 725
 QY 323 CGTCATGACTACATCA-----GGGAC-----CGT-----CA 349
 Db 726 SGDCYWNSSGCVDKTCANITLTSHASCYSIFNOCTVNNGGTACQTLATACTSYTQENCK 785
 QY 350 GTTGGAGC---TGAACCTCGGAATCGAAGTTGGACCCCGC-----GGGGCC 393
 Db 786 FTSNKVCVWTLACR---NATCADAPDITAYDSDETECLAYPTPSTCTVTVKVGAGCV 842
 QY 394 GAGGGC-----TGCGGCTCCGGGCTCC---415
 Db 843 SKSANCSDYMTSAQCHKTLNLTANDCKWIVDRCVALSFPATGAC---TTFKGNKTMCBG 900
 QY 416 ---GCT---CAGCACCTCAACGGCGAGATCAGC---GCCCTGAGC---CG 456
 Db 901 YRAGCTNVTGAASSASCTLDCTLKT-----GSLGTFADQALDSTCVSKDGTGCVIQST 956
 QY 457 AGCGGCGATCGTTCTCGCGAGCATCGC-----ATCTTGTTGTCG-----CT 498
 Db 957 CAGVGSFATNCFRSSASGTAGYCAMNTNCSVTSAAECAPVTGLTGLDHSKQLYHSSCT 1016
 QY 499 GAA-GGCG---CTCCCCCAGGAGCCGGCAGC---CC-----CA-----GCCATCCAG 539
 Db 1017 SLKDGTCGBYKACSSYATGNTCANSVQKGFDDATDCLRFANCASITGTGLTNTICTV 1076
 QY 540 GGGCGAAGAGAAATTAG-----TGCTCTGTGGGTC-----TCCC-----C 575
 Db 1077 YDPCVANVG---TACQELKATCAAVLTQNSCSTSTAGTCAMSGSACLTVVDANVATSC 1133
 QY 576 CAACGCGCCTCGCGC---GATCTGAGGGAG-----AACAGACCGATCG---CGGCGCATCGC 628
 Db 1134 AYITGTGLTNAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCSQSDAGLCWSSGA 1193
 QY 629 C-----CCTTAACCTGC-----ATCC 643
 Db 1194 CLTVVDANVATECPYITGTGLTNAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCS 1253
 QY 644 ---AGCCT-GGGGCTGAGGCTGAGGCACTCGCGAGAGAGGGCGCTCCTCTCTGCAC--- 696
 Db 1254 QSDAGLCWSSGACLTVVDANVATECPYITGTGLTDAICAG---YNAKCTVNRAGTACQK 1311
 QY 697 -----ACCTACTAGTACCAGAGACTTTAGGGGTGGGATTCACCTCGTGTGTT 745
 Db 1312 EALCATYAAVQATCSQSDAGLCW---SGSACLTV-----DANVATECPYITGTGTT 1361
 QY 746 T-----CTATTT-----TTTGAAGAAGC---AGACA-----T 768
 Db 1362 NAICAGYNAKCTVNRAGTACQKKEALCATYAAVQATCSQSDAGLCWSSGACLTVVDANV 1421
 QY 769 TTTAAAAATGGT---CA-----CGTTGGTGC-----TTC-----TCAGATT----- 803
 Db 1422 ATECAIYITGLTDAICAGYNAKCTNLKDGTCODEKATCKLVTQNKCTSQITGFLSCL 1481
 QY 804 -----TCTG-----AGGAATTTGCTT-----TGTAATTG 826
 Db 1482 WFDNSCSPIITDVTCISAIVQSLDHAQCAQAYSTGCTSVSDGSKQDFKTTCEQYAGTALSC 1541
 QY 827 TATATTACATGATC-ACCGACTGAGAAATATTGTTT-----TACA----- 865

Db 1542 TKIATSKCYLOGNCITISNVATDCAKITSGAGTITVEICQSYNTGCVNRSACVQQQ 1601
 QY 866 -----ATAGTCTGTGGCGCTGTTTTTTTGTATTATAACAAATAATTAGATGTA 919
 Db 1602 AQCSGYTSAMTSCYKSGAGLCIASTNTDT-----ACVAATAATTCDAVILGTGNY 1652
 QY 920 AA 921
 Db 1653 SA 1654

RESULT 7

Y278 MYCTU STANDARD; PRT; 957 AA.
 AC P56877;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PE-PGRS family protein Rv0278c/MT0291 precursor.
 GN Rv0278C OR MT0291 OR MV035.06C.
 OS Mycobacterium tuberculosis
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."; Nature 393:537-544(1998).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CDC 1551 / Oshkosh;
 RC MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."; J. Bacteriol. 184:5479-5490(2002).
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
 CC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcements>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AL021930; CAAL7353.1;
 DR EMBL; AE006936; AAK44511.1; ALT_INIT.
 DR PIR; D70835; D70835.
 DR TIGR; MT0291;
 DR TuberculList; Rv0278c;
 DR InterPro; IPR000084; PE_region.
 DR Pfam; PF00934; PE; 1.
 KW Hypothetical protein; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 957 HYPOTHETICAL PE-PGRS FAMILY PROTEIN

1238 KAGTACQKATCNLYTTEATCTSAATAATADKCAWSGAACLAATVTTATECAVVTGTLT 1297
 259 GAGCAGGTGAGATTCTC-CAGCAC---GTC-----ATCCACTACAT---CAGGCAC 344
 1298 NAICAAYN-----ANCTANKAGTACQKATCNLYTTEATCTSAATAATADKCAWSGA 1351
 345 C-----TTCAGTTG-----GAGCT-----GAAC-----TCGGA 368
 1352 CLAATVTTATECAVVTGTLTNAICAAVYNTANKAGTACQKATCNLYTTEATCTSA 1411
 369 TCCGAAGTGGGACG-----CCGGGG-----CCGAGGG-----TGC----- 402
 1412 AAATADKCAWSGAACLAATVTTATECAVVTGTLTNAICAAVYNTANKAGTACQK 1471
 403 CGTCCGGGCTCCGCTCAGC-----ACC-----CTCAACGGCG-AGATCAGCGCCCTGACG 452
 1472 CKDYTTSNKTAQTSTLSLWIDNSCVPTDLNCSVITGLGFVHAQCAQAYSGCTSVSD 1531
 453 GCGGAGGC-----GGCATGGTCTCTGC-----GGAGATCGGAT-CTTGTCGCG-- 496
 1532 G-----SKQDFKSTCEQVPTGLCTKASTKVLQSAKIT-SNVAITDCAKITSGAGTI 1587
 497 -----CTG-AAGCGCTCCCGCAGGACGGCGGACCC-----AGCC-----ATCAGG 540
 1588 TFEICQSYNTGCVNARSACVQQAQCGSYTSAMTSCYKSGAGLCIATNTTDTCAVAAT 1647
 541 GGGCAA-----GAGGAATTAG--TGCTCTGTGGTCTCCCGCAGCGCTCCCGGA-T 593
 1648 RAATCAVILGAGNYSANENKAGCTNGT--TACVAKTCAVAGITFNHNCNLYNT 1705
 594 CTGAGGGAGAACAGACGATCGGC-----GGCCACTCGC-----CCTTAATCGATCCAGC 646
 1706 CTVNSGNSACQTMASKADQTAQSLYSVEGECVVGTSVVRKTCDTAATDRDDTTC 1765
 647 CTGGGGCTGA-GGCTGAGGCA-----CTGG-----CGAGGA-- 676
 1766 STYQGSCHVALGCAQAAACATYKSLCKFNTSGKCFWNTKTCVDLNCNIEAT 1825
 677 -GAGGGCGCTCTCTCT------GC-----ACA-----CCTACTAGTTC----- 707
 1826 LYDTHNECVAVDNLACTVTRATNGAAGCGWARGACASYTIEQCKTNASNGVCVWNTNA 1885
 708 -----AC-----CAGAGACTTT-----AGGGGGTGGGATTCAC 736
 1886 NLPAFACQDKSCTSAPTSTTTNDYAYNTATVKCTVATPSNSGNPTLGGCQQTAA 1945
 737 -----TCGTGTGT-----TTCTA-----TTTTTGA 760
 1946 SSIYDKQQCINANGPCGWMGTQCADKSCATASATADYDDTKCRAYITNKCTVSDSQ 2005
 761 GC-----AG-----ACATTT--TAAATAATGGTCT--ACGTTT 788
 2006 GCVELPATCEMTQKQCYNKGADPCYWTGTACITKSCDNPATATADENYLAGTTL 2065
 789 GGTGCG-----TTCTCAGATTCTGAGGAATGCTTTGTATGTATATTAAT 837
 2066 NNVKCKTKVCEDFAPADLCKQAISCTTTNG-----TNCVTRGTCPQALSGACVTSST 2120
 838 GATCACCGA-----CTCAGAAATATTTTACAAATAGTTCTGTGGGGC-TGT 883
 2121 NQCEWIPAVLNASNVITSPAYCTIKMGSTAPILTSEACAGVFNCTTKGCGCVTKS 2180
 884 T-----TTTTGT-----TATTAACAATAATTATGATCG 914
 2181 TCSAVTIDVACTALNGTVCAWDSQNKCRDKOCQDFSGTTHAACQA-----ORAGCTAG 2235
 915 TGA 920
 2236 AGGKCA 2241

RESULT 6

G168 PARPR STANDARD; PRT; 2704 AA.
 AC P17053;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE G surface protein, allelic form 168 precursor.
 GN 1688.
 OS Paramesicium primaurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Paramesicium.
 OX NCBI_TaxID=5886;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=90172419; PubMed=2308165;
 RA Prat A.;
 RT "Conserved sequences flank variable tandem repeats in two alleles of
 the G surface protein of Paramesicium primaurelia.";
 RL J. Mol. Biol. 211:521-535(1990).
 CC -!- FUNCTION: This protein is the surface antigen or immobilization
 antigen of Paramesicium primaurelia.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DOMAIN: It has internal homologies and a highly periodic structure
 with 37 periods of about 75 residues, each period containing 8
 cysteines, except for four half periods. A variable part of 475
 residues comprises 4 almost identical periods in the middle of the
 protein.
 CC -!- MISCELLANEOUS: Expression of G protein occurs at low temperatures
 (14-32 degrees Celsius).
 CC -!- SIMILARITY: Contains 34 PSA repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X52133; CAA36378.1; -;
 CC FIR; S09118; S09118.
 CC InterPro; IPR002895; Paramesicium SA.
 CC Pfam; PF01508; Paramesicium SA; 34.
 CC SMART; SM00639; PSA; 33.
 CC Signal; Repeat; Antigen; Membrane; GPI-anchor.
 CC SIGNAL 1 20 POTENTIAL
 CC CHAIN 21 2704 G SURFACE PROTEIN, ALLELIC FORM 168.
 CC DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.
 CC FT DOMAIN 1060 1424 VARIABLE DOMAIN, COMPRISES 4 ALMOST
 CC FT IDENTICAL REPEATS.
 CC SQ SEQUENCE 2704 AA; 278775 MW; 40EA0A0B18EE2119 CRC64;
 Query Match 11.5%; Score 665; DE 1; Length 2704;
 Best Local Similarity 23.9%; Pred. No. 1.5e-26;
 Matches 344; Conservative 43; Mismatches 487; Indels 568; Gaps 77;
 QY 3 GGCCCATCTGTTTTCAGCCAGTCG--CCAAGAA--TCA-----TGAAAG-TCGCC 47
 Db 258 GGCVTTRTCAATTAATQASCIKNSGGDCYWTGTACVDTKTCANAPTMTTNSACAGFTGCI 317
 QY 48 AGTGGCAGCACCCGCCCGCGCC-----GCGGCGCC-----CAG-----CTGC 86
 Db 318 TKSGG--GCVANGACSVANVQAACVKNSSNFDCTTCKEKTCAAPTNNTHDLCTSY 375
 QY 87 CGCTGAAGCCCGGAGACA-----GGAGCGGTGC-----GGCGAGGT----- 127
 Db 376 LSTCTVKSQ--GCGCQNRSCANAPTMTTNDACEAYLTGNNCITKSGGCVTTTCAIT 432
 QY 128 -----GGTGC-----GC-----TCTCTCTCTGAGCAGA 150
 Db 433 LBAACVKNSSGCTCFWDTASSCKDKTCVNPATNTHTDLCOAFLNTCTVNSTSAGCCEK 492


```

RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grodin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RA "the complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to a
CC frameshift in position 85.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z95890; CAB09322.1; -.
DR EMBL; AE007040; -; NOT ANNOTATED CDS.
DR EMBL; BX248340; CAD94491.1; ALT_FRAME.
DR PIR; H70987; H70987.
DR TIGR; MT1807; -.
DR Tuberculist; Rv1759c; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR Antigen; Repeat; Signal; Complete proteome.
KW SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 914 WAG22 ANTIGEN.
FT SEQUENCE 91-4 AA; 74354 MW; F6953CDBE8B6AC8 CRC64;
SQ
Query Match 12.5%; Score 721.5; DB 1; Length 914;
Best Local Similarity 29.8%; Pred. No. 1e-29;
Matches 308; Conservative 20; Mismatches 377; Indels 329; Gaps 45;

```

```

468 ---GTTCTCTGCGACGATCGCATCTTGTGTGCTGAAGCGCTCTCCCGCCAGGA---C 518
534 NGDSGTGTGDDGAGGAGGWLFGNGG-NGAGAGAGTNGSAG-----GAGGAGGILFG 582
519 CGGCGGACCCAGCCATCCAGGGGCG-----AAGAGNATTACGTCTCTCTGG 567
583 TGGAGG-----AGGVGTAGAGGAGGAGGSAFLGSGGTGGVGAATTGGVG-----GAGG 633
568 ---GTCTCCCAACGCGCTCG-CCGGATCTG-----AGGAGAAACAAGACCGAT--CG 616
634 NAGLLIAGLGGCGGGAFTAGVTGTGAGGTGGAGLFGAGGAGGAGGTGTAGGAGGAG 693
617 CGGCGCACTGCGCCCTTAACTGCATCCAGCCCTGGGCTGAGGCTGAGGCACTGCG-GAGG 675
694 GAGGLYAHGG-----TGSPGNGSGTGTAGGTGGAGG-----PGGLYAGG 733
676 AGAGGGCGCTCTCTCTGACACACTACTAGTCACAGAGACTTTAGGGGGTGGGATTCCA 735
734 SGGAGHG-----GMAAGGGGAGGNA----- 754
736 CTCGTGTCTTCTATTATTTTGAAGAGAGACATTTTAAATAATGTCACGTTGCTGCTT 795
755 -----GSLTLNAGGAGGSGGSLSGKAGA-----GGAG--- 783
796 CTCGATTTCTGAGGAAATTCCTTTGATTATGATTATTAATGATCACCGACTGAGATA 855
784 ----GSAGLFYGGGAGGNGYSLNGTGGD-----GTGGAGQITGLRSGFGGAG 830
856 TTCTGTTTCAATAG-----TTCTGCGGCTGTTTTTTTATTATAAATAATTTAGA 911
831 AGGASDTGAGNGGAGGAGGAGLYGNGDGGAG-----GDGATSGKGGAGNAVVGNG 883
912 TGGTGAATAAAAAA 925
884 NGGNAGKAGGTAGA 897

RESULT 4
LORI_MOUSE STANDARD; PRT; 481 AA.
ID LORI_MOUSE
AC P18185;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Loricrin.
GN LOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90275605; PubMed=2190691;
RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
RA Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinart P.M.,
RA Yuspa S.H., Roop D.R.;
RA "Identification of a major keratinocyte cell envelope protein,
RT loricrin."
RL Cell 61:1103-1112 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/c;
RX MEDLINE=95256248; PubMed=7738016;
RA DiSepio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
RA Roop D.R.;
RA "The proximal promoter of the mouse loricrin gene contains a
RT functional Ap-1 element and directs keratinocyte-specific but not
RT differentiation-specific expression."
RL J. Biol. Chem. 270:10792-10799 (1995).
CC -!- FUNCTION: Major keratinocyte cell envelope protein.
CC -!- SUBUNIT: Monomers are crosslinked by disulfide and N-(gamma-
CC glutamyl) lysine isodipeptide bonds.
CC -----

```

RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AL022022; CAAL1745.1; --
DR PIR; F70806;
DR TubercuList; RV3508; --
DR InterPro; IPR000084; PE region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE region; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
Query Match 14.6%; Score 841; DB 1; Length 1901;
Best Local Similarity 32.3%; Pred. No. 2.1e-35;
Matches 335; Conservative 13; Mismatches 463; Indels 226; Gaps 40;
Qy 1 GGGGCCCATCTTCTTTTTCAGCCAGTCGCCAAGATCATGAAGTCGCCAGTCGAGCACCG 60
Db 180 GAGGWLFGVGGAGGVGGAG--GGTGGAGGPGGLTWGGGGAGGVGGGGTGGAGGGA--E 235
Qy 61 CCACCGCGCGCGCGCGCCAGCTGCGCTGAAGCCGCGCAAGACAGCG--AGCGGTGC 118
Db 236 LLFAGAGAGAGTGG-----PGATG-GTGGHGGVGGGGWLAFGAGAGGCGGAGAGS 290
Qy 119 GGGC--GAGGTGGT-----GGCTGTC---TGTCTGAGCAGAGCGTGCCCATCTC-CGCT 168
Db 291 DGGALGTGGTGGTGGAGGAGGRCALLLHAGGGGGLGAGCGGQGTGGAGGDLVGGVGT 350
Qy 169 GCGCG-GCGCGCGGGCGG--CCTGCTGCTGCTGCGAGCAGCAGCGTAAAGTGC 225
Db 351 GKGGGVGVAGLGGAGGAAGQLFSAGGAAGVGGTGGGGGAGGAGAGADAPA-STGL 409
Qy 226 TGCTCTACGACATG-----AACGCTGTCTACTCACGCTCAAGAGAGCTGTGCCACCT 280
Db 410 TGGTGFAGGAGGVGGCGNATAGG-----SGAGGTGGGAGGAGGGS 455
Qy 281 GCGCCAGAACCGAAGGTGACAGGTGGAGATTCTCCAGCAGTCACTACATCAG 340
Db 456 GADNAGTIGADG-GAGGTGAGAGGAGGAAAGT-----GGTGGVGAAGKAGIGTGGQ 508
Qy 341 GGA-----CCTTCAGTTGAGCTGAATCGGAATCCGAAGTTG-----GGACCCCGGG 389
Db 509 GGAGGAGAGTADATATGATGTTGTFSGAGGAGGAGGNTGVGTNGSGGQGT-----GGA 563
Qy 390 GCGCGAGS-----GCTGCCGCTCGGGCTCGCTCGACACCTCAACGCGGAGATC 440
Db 564 GGAGGAGGVADNPTGTCGTGGTGGKAGAGAGG-----GSSGAGGTTN 608
Qy 441 AGCGCCTGACGCGCGAGGCG-----GCATGCTTCTCGCGAGCA-TGCGATCTTGT 492
Db 609 GSGAGGTGGCGAGGAGAGAGADNPTGIGAGGTG-----GTGAGAGAGAGAGTGGT 663
Qy 493 -GTGCTGAAGCGCTCCCCAGGAGCCGCGGA-CCCCAGCATCCAGGGGCAAGAGG 550
Db 664 GGAAGVSVNAGIG-----GTGCTGGVGGAGGAGAGAAAGSSAT-----GCAGFAGAGG 712
Qy 551 AATTACGTGCTCTGGGTCTCCCGCCACGCGCTCCCGGATCTGAGGAGAACAGAC 610
Db 713 EGGAGGNSGVGTNGSG-----AGGAGGKGGTGGAGSGADNPTGAG 755

Qy 611 CGATCGCGCGCCACTGCGCCCTTAACTCG--ATCCAGC-----CTGGGGCT----- 654
Db 756 FAGGAGGTGGAGAGGAGGATGCTGCTGCGVWGTAGTSGAGTGGGGGCGDGGASGLGLG 815
Qy 655 ---GAGGCTCAGGCACATGGC-GAGGAGAGGCGCTCTCTCTGCACACCTACTAGTCAC 709
Db 816 GFDGGGCGGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 864
Qy 710 CAGACATTTTAGGGGCTGGGATTCACATCGCTGCTGCTTATTTTGTAAAGCAG----- 764
Db 865 VGGDGGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 915
Qy 765 -----ACATTTTAAATAATGTCACGTTTGTGCTTCTCAGATTT----- 804
Db 916 GCGGAGAGGCGGGGGLGAGATTTTINAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 975
Qy 805 -----CTGAGCAAAAT-----GCTTTGATTTGATATATTACAAATGATCACC 844
Db 976 DAGSGGGGFGGAGAGKAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1034
Qy 845 GACTCAGATATTTGTTTACATAGTCTG-----TGGGCTGTTTTTTTGTATTAAAC 899
Db 1035 GGINAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1094
Qy 900 AAATAATTTAGATGCTG 916
Db 1095 AAGGCGGAGGAGGAGGAG 1111
RESULT 3
WA22 MYCTU STANDARD; PRT; 914 AA.
ID WA22 MYCTU AC O06794;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WAG22 antigen precursor.
GN WAG22 OR WAG22B OR RV1759C OR MT1807 OR MTCY28.25C OR MB1789C.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.B., Taylor K., Whitehead S., Barrell B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
SEQUENCE FROM N.A.
RP SPECIES=M.bovis; STRAIN=AF2122/97;
RC

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:11:40 ; Search time 17 Seconds
(without alignments)
2836.292 Million cell updates/sec

Title: X77956
Perfect score: 5766
Sequence: 1 GGGGCCCATCTGTTTCAGC.....TTAGATGGTGAATAAAAAA 926

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	855.5	14.8	5263	1 FBOH_BOMMO	P05790 bombyx mori
2	841	14.6	1901	1 YZ08_MYCTU	O53553 mycobacteri
3	721.5	12.5	914	1 WAZ2_MYCTU	O06794 mycobacteri
4	706	12.2	481	1 LORI_MOUSE	P18165 mus musculu
5	680.5	11.8	2715	1 G156_PAPPR	P13837 paramescium
6	665	11.5	2704	1 G168_PAPPR	P17053 paramescium
7	647.5	11.2	957	1 Y278_MYCTU	P56877 mycobacteri
8	635.5	11.0	747	1 SPD1_NEPCL	P19837 nephila cia
9	629	10.9	778	1 YQ34_MYCTU	P71333 mycobacteri
10	627	10.9	801	1 Y747_MYCTU	O53810 mycobacteri
11	618.5	10.7	672	1 PHX5_MOUSE	P08399 mus musculu
12	570.5	9.9	2321	1 NTC3_HUMAN	Q9um47 homo sapien
13	562.5	9.8	2319	1 NTC3_RAT	Q9r172 rattus norv
14	554	9.6	2531	1 NTC1_RAT	Q07008 rattus norv
15	553	9.6	2531	1 NTC1_MOUSE	Q01705 mus musculu
16	551.5	9.6	2318	1 NTC3_MOUSE	Q61982 mus musculu
17	547.5	9.5	4289	1 TENX_HUMAN	P22105 homo sapien
18	544.5	9.4	2556	1 NTC1_HUMAN	P46531 homo sapien
19	543	9.4	1700	1 BAR3_CHITE	Q03376 chironomus
20	527.5	9.1	641	1 EBNI_EBV	P03211 Epstein-bar
21	518	9.0	2003	1 NTC4_HUMAN	Q99466 homo sapien
22	517.5	9.0	1964	1 NTC4_MOUSE	P31695 mus musculu
23	510	8.8	1046	1 PSTA_DICDI	P11976 dictyosteli
24	509	8.8	1064	1 FEPI_STRPU	P10079 strongyloce
25	502.5	8.7	2437	1 NTC1_BRAP	P46530 brachydanio
26	501.5	8.7	5376	1 ZAN_MOUSE	O88799 mus musculu
27	496	8.6	2524	1 NTCX_XENLA	P211783 xenopus lae
28	493.5	8.5	465	1 GRP2_PHAVU	P10496 phaseolus v
29	482.5	8.4	2471	1 NTC2_RAT	Q9qk30 rattus norv
30	486.5	8.4	603	1 YD25_MYCTU	Q10637 mycobacteri
31	484.5	8.4	2471	1 NTC2_HUMAN	Q04721 homo sapien
32	481	8.3	384	1 GRP1_PETHY	P09789 petunia hyb
33	478	8.3	2470	1 NTC2_MOUSE	O35516 mus musculu

ALIGNMENTS

RESULT 1

FBOH_BOMMO STANDARD; PRT; 5263 AA.
AC P05790; Q17220; Q26379;
DT 01-NOV-1988 (Rel. 09, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).
GN FIBH.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330362; PubMed=10871375;
RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Enault C.,
RT Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;
RA "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL Nucleic Acids Res. 28:2413-2419(2000).
RN [2]
RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
flanking, mRNA coding, entire intervening and fibroin protein coding
regions.";
RL Cell 18:591-600(1979).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
RT "Structural analysis of the fibroin gene at the 5' end and its
surrounding regions.";
RL Cell 16:425-436(1979).
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX STRAIN=Kishu X Showa;
RC MEDLINE=89094868; PubMed=3210244;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
structure of silk fibroin mRNA.";
RL J. Mol. Biol. 203:917-925(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
gene.";
RL J. Mol. Evol. 38:583-592(1994).
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RC STRAIN=J-139;
RX MEDLINE=99296390; PubMed=10366732;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,

P07207 drosophila
P54320 mus musculu
P27483 arabidopsis
P12021 sus scrofa
P23490 homo sapien
P26372 ovio aries
O75690 homo sapien
Q50615 mycobacteri
Q99372 rattus norv
P20730 bombyx mori
Q61555 mus musculu

34 465 8.1 2703 1 NOTC_DROME
35 452 7.8 860 1 ELS_MOUSE
36 450.5 7.8 349 1 GRP_ARATH
37 445 7.7 1150 1 APMU_PIG
38 431 7.5 316 1 LORI_HUMAN
39 426 7.4 182 1 KRUC_SHEEP
40 422.5 7.3 194 1 KRUB_HUMAN
41 418 7.2 498 1 YIL8_MYCTU
42 416.5 7.2 713 1 TSA4_GIALA
43 414 7.2 864 1 ELS_RAT
44 407 7.1 178 1 CHHC_BOMMO
45 403.5 7.0 2907 1 FBN2_MOUSE

C:Superfamily: collagen alpha 1(IIV) chain

Query Match 11.9%; Score 686.5; DB 2; Length 1660;
Best Local Similarity 29.8%; Pred. No. 1e-24;
Matches 305; Conservative 20; Mismatches 460; Indels 239; Gaps 39;

QY 1 GGGGCCCATCTCTTTTACGACAGTCCCAAGAA--TCATGAAGTCGCCAGTGGCAGCAC 58
DB 141 GNG-----AGAVGVGAGAGAGLFGIGAGGAGGAGAPGGTGTGWLGGGGVGGNG- 196
QY 59 CGCCACCGCGCGCGGCCCGCCAGTGCCTGTAAGCGCGGAGAGACAGACGAGCGGTGC 118
DB 197 -----GAGG-----GAGGAGGAGNAGLFGNGGAGGAGGAGGAGGAGG 232
QY 119 GGGC-----GAGGNGG-CGCTGCTGCTGTGACGAGCGTGGCCATCTCGCGTCCCGG 173
DB 233 NAGWFGHGGAGGAGGAGGAGAGATGCGDGAAGVAGSDGAG-----GGLAGSDGG 284
QY 174 -GGCGCGCGCGCGCGCTGCTGCTGCTGTCGACGAGCGTGGACGAGGTAAAGTGTCTG-CTCT 231
DB 285 DGGAGGVGNG-GRG--GWLLGNGGAGGVGGVGGAGGAGAGGAGGAGATGNGPAGIS 340
QY 232 ACACATGAACGGCTGTACTACGCTCAAGGAGTGGTGGCCACCTCTGCCCGCAGAAC 291
DB 341 AAGGDDGAGNGGAGGNG-----GVGAGGAGGS-----AGLGYVGRAGD 381
QY 292 GCAAGGTGAGCAA--GGTGGAGATTTCTCCAGCAGCTCATCGACTACATCAGGACCTTCA 349
DB 382 GGAGGGGELGAPDGGGAGGNGSWLAAGDGGAGGHPGLGA-----GGA-----G 430
QY 350 GTTGGAGCTGAA--CTCGGAATCGAAGTTGG--GACCCCGCGGGCGGAGGCTGCCGT 406
DB 431 GASGAGARAGANGLAAGNDGPGVSGNGGKGNHAPVAGHGGNGGAGGN-----GGL 485
QY 407 CCGGGCTCCGCTCAGCACCCCTCAACGCCGAGATCAG-----CGCCCT-----GACGGC 454
DB 486 VDDG-----AGHGGGAGAGGYADMTAIFLSSGTPGSDGNGGA 527
QY 455 CGAGGCGGCATCGTTCTCGGACGATCG-----CATTTGTGCTGCTGAAGCCCTC 508
DB 528 GGAGGAGAGAGGAGGAGGAGGAGGAGGAGFNAVLVSDGNGDGGAGGRG-----583
QY 509 CCCAGGACCGCGGACCCCGACCATCCAGGGGGCAAGAGGAATTACGTGCTCTGTGG- 567
DB 584 -----GDGA--GGAGD--APAGAGGQGVGGDGG--AGGAGAP-----GNG 626
QY 568 GTTCTCCCGCAACCGCGCTCGCGGATCTGAGGAGAGAAAGAGCCGAT-----CGCGGCC 622
DB 627 GD-----NAFKDGDGAGDGDGDPGAGGKGGAGGAGATGVTGATGATVHSGNGG-- 677
QY 623 ACTGGCCCTTAACCTGCATCCAGCTCGGCTGAGCTGAGGCAC-----667
DB 678 -KGGNGADATVAGANG-----GKGGAGNGGLVGGAGGDDGGGAGAGANAVGEDGADG 732
QY 668 -----TGCGAGGAGAGGGCGCTCTCTCTCTGACACCTACTAGTAC 709
DB 733 TLGGOPGEGSEANGGQGGVGGGAGGAGGAGGAGSSALGSGNGGRDAGAGGAGGAGG 792
QY 710 CAGAGACTTTAGGGGGTGGGATTCATCTGCTGCTTTCTATTTTTGA 760
DB 793 AGGAGGVSQGGFPGKGGAGGAGAGAGAGGGGGGKAGSADSAEAVGAGGKGGDGGVGG 852
QY 761 -----CGAGACATTTTAAA-----AAATCGTCACGTTTGGTCTTCTCAGATTTT 805
DB 853 VGGDGGPGDGGAGGAAPAGQVGSVGVGVDGGLGAGGNGDGGHSDGGDGGDGG 912
QY 806 TGAGGAATTTGTTTGTATTGTATATACATGATCACCAGTACAGAAATTTTGTATACA 865
DB 913 PGAGGLGLGDSNGNTR-----AASGVDSADHDPGSGGNGGNGGNGAGAAQAVAG 961
QY 866 ATAGTCTG-----TGGGGCTGTTTTTTTGTATTATAACAAATTAATTTAGATGCTGAAA 920
DB 962 GAGNGGDDGNAGRVGDGAGGNGDGAAGANGANSAGPSDALALQPGNGGQGDAGQ 1021

QY 921 AAAA 924
DB 1022 AGGA 1025
RESULT 15
S09118
G surface protein 168 - Paramesium primaurelia
C:Species: Paramesium primaurelia
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 07-Dec-1999
C:Accession: S09118
R:Prat, A.
J. Mol. Biol. 211, 521-535, 1990
A:Title: Conserved sequences flank variable tandem repeats in two alleles of the G sur
A:Reference number: S09118; MUID:90172419; PMID:2308165
A:Accession: S09118
A:Molecule type: DNA
A:Residues: 1-2704 <PRA>
A:Cross-references: EMBL:X52133; NID:gl0049; PIDN:CAA36378.1; PID:g578473
C:Genetics:
A:Genetic code: SGC5
C:Superfamily: G surface protein
Query Match 11.5%; Score 665; DB 2; Length 2704;
Best Local Similarity 23.9%; Pred. No. 1.3e-23;
Matches 344; Conservative 43; Mismatches 487; Indels 568; Gaps 77;
QY 3 GGCCCATCTCTTTTCAGCCAGTGC--CAAGAA-----TCA-----TGAAG-TGCC 47
DB 258 GGCVTRTTCAAAATQASCINKNSGGDCVWTGTACVKTCCANAPTMTTNSACAGFVTGCI 317
QY 48 AGTGGCAGCCGCCACCGCGCC-----GCGGGCC-----CAG-----CTGC 86
DB 318 TKSGG--GCVANGACSVANVOAA CVKNSNPDCITWDTTCKEKTCCANAPTMTTNTDLCTSY 375
QY 87 CGCTGAAGGCGGCAAGACA-----GCAGCGGTGC-----GGCGAGGT----- 127
DB 376 LSTCTVKS-----GGCQNRSCANAPTTMTTNDACEAYLTGNNCITKSGGCVTNTTCAIT 432
QY 128 -----GGTGC-----GC-----TGCTGCTCAGCAGA 150
DB 433 LEAACVKNSSGTCFWDTPASSCKDKTCVNAPATNTTDLCOAFLNTCTVNSTSAGCCEK 492
QY 151 CGCTGGCCATCTCGCGCTGCGCGGC-----GCGGGGCGCGCTGCTGCTGCTGCTGCTG 203
DB 493 TCENSLVLAICDKDTSSBACIWKGYKQCVLAS-----SATTTHADCTVHSTCTL 545
QY 204 GACGAGCA-----GCGAGT-AAACGTGCTGCTCTAGCAGATGAACGGC--TGTTACTCAG 256
DB 546 SNSGTGCVPLPLKCEAITIEAACNLKANGPCGWNGSQCIDKACSTASKTFTTTSQCTGH 605
QY 257 CCTCAAGGAGCTGGT-----GC-----CCACCTGCC-----CCAGAA 289
DB 606 ISTCVANNPVTNGSLTIGQCDLPTSCAARKSENCEIARVGPPTCLWVSSSTSCVRKS 665
QY 290 CGCAAGGTGAGCAAGTGGAGATTC--TCCA-----GC-----A 322
DB 666 CATASTVGTGTGALSAGGTGSGCTYLTNCISNNATDGCIAKPSSCSLVSNCRDGSKA 725
QY 323 CCGTCATCAGTACATCA-----GGGAC-----CTT-----CA 349
DB 726 SDCTVWNGSSCVKDTCANITLTHASCVYSIFNQCTVWNGGTACOTLATATCTSYSTOENCK 785
QY 350 GTTGGAGC--TGAATCGGAATCGCAAGTGGGACCCCG-----GGGGCC 393
DB 786 FSTNKNKCVMTGLACR---NATCADAPDPTTAYDSDECLAYPTPSETCTVYVYKVAOQCV 842
QY 394 GAGGGC-----TGCGGTCCCGGCTCC-- 415
DB 843 SKSANCSDYMTSAQCHKLTNLTANDCKWIVDRYALSSPATGAC--TTFKNTMTCEG 900
QY 416 ---GCT-----CAGCACCCCTCAACGGCAGATCAGC-----GCCCTGACG--GC-----CG 456

QY 871 TCTGTGGGCTGTTTCTTTTGTATTAAACAAATAATTTAGATGCTGTAAGAAAAA 924
Db 847 STLYGNGNGN-----GGAGGTGGKAGVGA 875

RESULT 13

A23475
G surface protein - Paramacium primaurelia
C:Species: Paramacium primaurelia
C:Date: 31-Mar-1999 #sequence_revision 31-Mar-1989 #text_change 07-Dec-1999
C:Accession: A23475
R:Prat, A.; Katinka, M.; Caron, F.; Meyer, E.
J. Mol. Biol. 189, 47-60, 1986
A:Title: Nucleotide sequence of the Paramacium primaurelia G surface protein. A huge pro
A:Reference number: A23475; UID:9706934; PMID:3783679
A:Accession: A23475
A:Molecule type: DNA
A:Residues: 1-2718 <PRA>
A:Note: the authors translated the codon TGC for residue 2665 as Trp
C:Genetics:
A:Genetic code: SGC5
C:Superfamily: G surface protein

Query Match 12.0%; Score 689.5; DB 2; Length 2718;
Best Local Similarity 25.8%; Pred. No. 1e-24;
Matches 358; Conservative 31; Mismatches 492; Indels 505; Gaps 82;
4 GCCATTCTGT-TTCAGC---CAGTCGCCAAGAAATCATGAAGTCGCCAGTGCACACC 59
892 GAC--TTFKJKTWCEYRAGCTNVG--AASSASC-----TLDTLKTG--SGLTEA 938
60 GCCACCGCGCGCGGCGCC---CAGCTCGCGGC-TGAAGCGCGGCAAG-----ACA 107
939 DCQALDSTCVYKDGTCIAIQSTCAGYGTAAACFRSSASGTAGYCAMMTCQSVTSAA 998
108 GCGAGCGGTG-----C-----GGCGAGGTGTGCGCTGTCT---GTC--- 142
999 ECAFVTGLTGLDHSKQLYHSSCTSLKDGTCQEKYKTCGXYAATNCAATGSGKCFPDV 1058
143 -----TGACGAGAGGT---GGCAT-----CTGCG-----GTCGCGGG 174
1059 ECLRFNSCASITGTGLTATIGTYDAGCVANVGTACQEKLATCDLYLTQNSCTSA 1118
175 GCGCGGGCGCGCTGCTGCGCC---TGCTGGACGAGCAGCA---GGTA-----AACGT 223
1119 TADKAMSGTACLAATVTVGTHCAVYTGTLTDLICAYNANCTANKAGTACQEKATCNL 1178
224 GCTGCTCTACACAT-----GAACGGCTGTACTCACGCT-----CAAGGAGCT-- 268
1179 YTTTATCTSAATAATADKAMSGAACLAAT--TVATECAVYTGTLTDLICAYNANCTAN 1237
269 -GGTGC-----CC-----ACCCTGC-----CCCGAAGCCG-----CA---AGGT 298
1238 KAGTACQEKATCNLYTTEATCTSAATAATADKAMSGAACLAATVTVATECAVYTGTLT 1297
299 GAGCAAGGTGGAGATTCTC-CAGCAC---GTC-----ATCGACTACAT---CAGGAC 344
1298 NAICAAYN-----ANCTANKAGTACQEKATCNLYTTEATCTSAATAATADKAMSGAA 1351
345 C-----TTGAGTTG-----GAGCT-----GAGC-----TCGGAA 368
1352 CLAVTVATECAVYTGTLTNAICAYNANCTANKAGTACQEKATCNLYTTEATCTSA 1411
369 TCCGAAGTTGGGACC-----CCCGGGG-----CCGAGGCG-----TGC----- 402
1412 AAATADKAMSGAACLAATVTVATECAVYTGTLTNAICATYNAGCINLKGDTGQCEAKAN 1471
403 CGTCCGGGCTCGCTCAGC-----ACC-----CTCAACGGCG-AGATCAGCGCCCTGACG 452
1472 CKDYTTSNKTAQTSTLSCLWIDNSCPVTDNLNCSVITGLGFVHAQCCQAYSTGCTSVSD 1531
453 GCGGAGGC-----GGCATGGCTTCCTGC---GGACGATCGCAT-CTTGTGTGCG-- 496

Db 1532 G-----SKQDPFKSTCEQYFGTTLGCTKTASTKCYLQSGACITISNVATDCAKITGSAGTI 1597
QY 497 -----CTG-AAGCGCTCCCGCAGGAGCGCGGACCC-----AGCC-----ATCCAGG 540
Db 1588 TFEICQSYNTGCSVNRARSACVQQAQCSGYTSMWTSYKSGAGLCIATINTDTACVAAT 1647
QY 541 GGGCAA---GAGGAATTACG---TGCTCTGTGGGTCTCCCGCAACCGCGCTCGCCGGA-T 593
1648 AATCAVYLAGNYSSANCMKAGCTNNGT--TACVAKTCAAAAGITFNHTNCNSYLT 1705
QY 594 CTGAGGAGAACAGACCGATCGC-----GGCCACTCGC-----CCTTAACATGCATCCAGC 646
1706 CTVNSGNSACOTMASKADQTOASCLYSEVGEVUVGTSCURKTCDDTAADTDDTEC 1765
QY 647 CTGGGCGTGA--GGCTGAGCA-----CTGG-----CGAGAA-- 676
1766 STYQSCCTVARLGACQARAACATYKSLQCKFNFTSGGKCFWNPNTKTCVDLNCNIBATT 1825
QY 677 ---GAGGCGCTCTCTCT-----GC-----ACA-----CCTACTAGTC----- 707
1826 LYDTHNECVAVDATLACTVVRATNGAAAGCGMARACASATYIBEOCKTNASNGVCVWNTNA 1885
QY 708 -----AC---CAGAGACTTT-----AGGGGGTGGGATCCAC 736
1886 NLPAPACQDKSCTSAPTSTTTHNDYAYNTATVKCTWATPNSGNGNPTLGGCQTAAAC 1945
QY 737 -----TCGTGTG-----TTCTA-----TTTTTGA 760
1946 SSIYDKBOCQINANGDPCGMNGTQCADKSCATATADYDDDTKCRAYITNKCTVSQ 2005
QY 761 GC-----ACATTT---TAAAAAATGGTC---ACGTTT 788
2006 GCVEIPATCETMTQKQYNNKAGDPYWTGTACITKSCDNAPDATATADECNLYAGCTL 2065
QY 789 GGTGC-----TTCTCAGATTTCTGAGGAAATGCTTTGTTATGTATATTAAT 837
2066 NNKCKTKVCEDFAPATDALCKQAISTCTTG-----TNCVTRGTCFQALSOAGCVTSST 2120
QY 838 GATCACCGA-----CTGAGAATATCTTTTACAATAGTCTGTGGGCGC-TGT 883
2121 NQCEWIPAVLNASNVITSPAYCTIKNCSTAPILTSEACAGCTTCTTNGGCGCVTKS 2180
QY 884 T-----TTTTGT-----TATTAAACAAATATTTAGATGG 914
2181 TCSAVTIDVACTTALNGTVCAMDSQNKCRDKCQDFSGTTHAACQA-----QTAGCTAG 2235
QY 915 TGAAAA 920
2236 AGGKCA 2241
Db

RESULT 14

A70869
Hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70869
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genom
A:Reference number: A70500; UID:98295987; PMID:9634230
A:Accession: A70869
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1660 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:G3261507; PIDN:CAAL6067.1; PID:G2791
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2490c

Db 361 LGLDANAPASTNPLHTAQOQALAAVNAPIQAVTGRPLIGNGANGAPSGAPFGHGWL 420
 Qy 441 AGCGCCCTGACCGCGGCGGCAATGCGTCTCTGCGGACGATCGCATTTTGTGCGTGA 500
 Db 421 GGGGTGGSGVSGAGGCGGAGILFGAG-----GAGGAGGA-----VTGTGATGSGG 468
 Qy 501 AGCGCTCCCGCAGGACCGGCGGACCCAGCCATCCAGGGGCGCAAGAGGAATACGTGC 560
 Db 469 AGGG-----ALLFGAGA-----SGIGFAAGGAGPG-----GAGG 506
 Qy 561 TGTGTGGGTCTCCCCAAAGCGCTGCGGATCTGAGGAGAAACAAGACCGATCGCGG 620
 Db 507 LFNNGGAG-----GAGGSGVSGAGGCGGAGGAGGILFAGGAGGAGG 548
 Qy 621 CCACTGGCCCTTAATCTGATCCAGCGCTGGGCTGAGCTGAGGCACTGGCG--AGGAGA 678
 Db 549 GGNVNG-----GAGGAGGVGGLFAGGAGGSGGSGVAGDSGA 586
 Qy 679 GGGCGCTCTCTCTGACACCTACTAGTCAACAGAGACTTTAGGGGGTGGGATCCATC 738
 Db 587 GGNAG-----LLAPLAGGAGGGGCGGDFDTGAGGPGGDA-----GLIV 625
 Qy 739 GTGTGTTTCTATTTTGAAGACACATTTTAAATGCTCACGTTTGTGCTTCTC 798
 Db 626 GSG-----GVGAGGFLTTGGPG-----AA 646
 Qy 799 AGATTTCTGAGGAAATGCTTTGTTATTTGATATTAACAATGATCACCAGTGAATATTG 858
 Db 647 GPDAGLLFGSGAGGAGGSGRTDLGGAGGAGGKAGLIGNGN-----CGAGGAGNGG 699
 Qy 859 TTTTACAAATGTTCTGTGGGCTGTTTTTTTGTATTATTAACAAATTTAGATGTT 915
 Db 700 GDGGPGGAFGLGNGGNGGNGTGT-----SAGSPGAGGAGGS 737

RESULT 10
 A70934
 hypohetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70934
 R: Cole, S. T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrall, B. G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: A70934
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-1306 <COL>
 A: Cross-references: GB: AL021942; GB: AL123456; NID: g3242298; PIDN: CAA17449.1; PID: g290963
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: Rv0578c
 C: Superfamily: collagen alpha 1(IV) chain

Query Match 12.3%; Score 712; DB 2; Length 1306;
 Best Local Similarity 30.2%; Pred. No. 6.4e-26;
 Matches 332; Conservative 20; Mismatches 464; Indels 282; Gaps 39;

Qy 1 GGGGCCCATCTCTTTTCAGCCAGTCGCGCAAGATCATGAAA--GTCCCACTGTCGACAC 58
 Db 167 GAGG-----VGTGGAG-----GHGAGGWLNGGAGGFGGAGVGVNGGAGGTA- 212
 Qy 59 GCGCACCGCCGCGCGG--GCCCCAGCTGC-----GGCTGAAGGCCGCGCAACACACGA 111
 Db 213 -----GLFGVGGAGGAGNGIAGVTGTSTASTPGSGTAGGAGGIGGNGGAGGAGGVLM 265
 Qy 112 GCGGTGCGGCGAGGTGTT--GGCTGTCTGTCTGACGACGAGCGTGGCCACTCTCG----- 164
 Db 266 GNGGN--GGAGGEGPGGAGAGAGAGATWLGADGAGGNGGNGAGGTGCGVGGPGGH 323

Qy 165 --CGCTCCCGGGGCGCGG-----GGCGCGCTCCCTGCCCTGCTGACCGAGCAGCAG 215
 Db 324 GLLGLGSHAGAGAGAGGSGGPGGNGATGTWGNHLAGGTGNGNGNPGAGGAGGAG 383
 Qy 216 GTAAA CBTGCTGCTTACACATGAACCGCTGTACTACCGCTCAAGAGAGCTGTGCCCC 275
 Db 384 G--ASVGSASHAN--GAPGTTSTSGNGDGGKADAISSG--QTGANGRGSGDGG-- 435
 Qy 276 ACCTCCCCCAGAACCGAAGGTGAGC-----AAGGTGGAGATTCTCCAGCACGTCT 328
 Db 436 ----VNGGAGGAGRG--GAGGLGFSBAPRPGAGGTGGAGN-----GGTQAGDGGT 485
 Qy 329 CGACTACATCAGGACCTTTCAGTTGAGCTGAATCCGAATCCGAATTTGGGACCCCCGG 388
 Db 486 GGA-----GGA-----GGDGGSGGAGSIFGNASAPGAAGSPGNGNGGPGG 527
 Qy 389 GGGCCGAG-----GCTGCCCTCCGGCTCCGCT-----CAGCA 423
 Db 528 AGGEGAGGALAAASGONGSQAGGCGAGNGGCTPENGHGAAGALGVNGVGGAGG 587
 Qy 424 CCCTCAACCGCGAGATCAGCGCC--CTGACCGCGAGGCGGCGATCGTTCTCT----- 473
 Db 588 GPPGVGAGGCGGSGSTPGANGAPGNTPTSGNGGNGRGADATGFGGTGASGGRGDDG 647
 Qy 474 --GGGACGATCGCATCTTGTCTGCTGAAGCGCTCCCCCGAGGACCGCGGACCCCG 531
 Db 648 LVNGGAGGAGGNGSKGLPLGLHGLNPLDG-----GTGNGGAGSGGA---WAG 695
 Qy 532 CCATCCAGGGGCGAAGAGGAATTAACGTCTCTGTGGCTCTCCCCCAACCGCCCTCGCCG 591
 Db 696 NGCTGGAGTGGVGTGCGSGDGVNGSSAGADHPGCT-----GGVGG 738
 Qy 592 ATCTGAGGAGAACAGACCGATCGCGCGCCACTGC-----GCCCTTAAGTCATC 642
 Db 739 TGGKGGDGGGAAPNGVAGSQGPGAGDGGTGGVNGGNGRIDGADGATAGARG--QDG 797
 Qy 643 CAGCTGGGCTGAGGCTGAGGCACTGGCGAGGAGAGGCG-----GCTCCTCTCTG 693
 Db 798 CAGGAGGKGRGTGGTGGGAGPAGTTGSGAGGNGSGGTGGDPGCGGANGSVFTNG 857
 Qy 694 CACACTACTAGTCACACAGAGACTT-----TAGGGGTGGATTCCACTCGT-----GTT 745
 Db 858 ICGNGGNGNAGPSGAGSGGAGSTFGATGSSSIHVNNGNGNGNDHALSNGAAGG 917
 Qy 719 -----TAGGGGTGGATTCCACTCGT-----GTT 745
 Db 918 NGNGNGSLRSGGAGHNGGNGNAGRMGDDGTGAGGAGGAGGAGGAGGAGGAG 977
 Qy 746 TCTATTTTGAAGACAGACATTTTAA-----AAATGTCACGTTTGGTCTT 795
 Db 978 GSDGNPGAITGSGCRGGDGGVGGGSGVAGDAGDGRGAGGTGGTGLRG--TTGATGATG 1036
 Qy 796 CTCAGATTCTGAGAAATGCTTTGTATTTATATATACATGAT---CACCCACTGAG 851
 Db 1037 TFDAGA-----DHGNGGTGVTGAGGGGNGGAGGAGGAGGAGGAGGAGGAG 1092
 Qy 852 AATATTGT-----TTTACAATAGTTCTGTGGGCTGTTTTTTTGTATTAAACAAAT 903
 Db 1093 GAGGTGGTGGDGRGAGHGTFFSLAGT--GGTGGNGGTG-----GTGTGGAGGAGGT 1143
 Qy 904 AATTAGATGCTGAAAA 921
 Db 1144 G--STLGATGATGAAGA 1159

RESULT 11
 A35628
 loricrin - mouse
 C: Species: Mus musculus (house mouse)
 C: Date: 21-Sep-1990 #sequence revision 21-Sep-1990 #text_change 13-Aug-1999
 C: Accession: A35628
 R: Mehrel, T.; Hohl, D.; Rothnagel, J. A.; Longley, M. A.; Burdman, D.; Cheng, C.; Lichti,

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-914 <COL>
A;Cross-references: GB:295890; GB:AL123456; NID:9324245; PIDN:CA809322.1; PID:92131027
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv1759c; wag22
C;Superfamily: elastin

Query Match 12.5%; Score 721.5; DB 2; Length 914;
Best Local Similarity 29.8%; Pred. No. 1.9e-26;
Matches 308; Conservative 20; Mismatches 377; Indels 329; Gaps 45;

QY 2 GGGCCCATCTGTTTC-----AGCCAGT-----GCCCAAGAATCATCAAAAGTCCCCAGTGG 52
DB 83 GGGAYAAEAATVPLINSINAPVLAATGRPLINGANGAP--GTGANGDAGMLINGNG 140
QY 53 CAGCACCGCCACCGCCCGCCCGCCAGCTCGGCGCTGAAGCCGCGCAACACGCA- 111
DB 141 AGG-----SCAKGANGAGGPGGAAGLFGNG-GAGAGGTATVANNIGGAGGAGS 190
QY 112 ---GCGGTGCGGCG-----GAGGTGGTGGCTGTCTGTCTGAGCAGAGCGTGGCCATC 161
DB 191 AMLFGAGGAGGAGATSLVGGTGGTGTG-GNAGML-----AGAAAG-GAGGSPST 242
QY 162 TCGCGCTGCGCG-----GGCCCGCG-----GGCGCGCTGCTGCTGCTG-----GACGA 208
DB 243 AGGAGGAGGAGGLFTTGGVGAGGCGGHTGGAGGAGGAGLFGAGMGAGGFGFDHGTLT 302
QY 209 GCAGCAGTAACTGTCTTACGACATGAACGGCTGT--TACTCAGCCTCAAGGAG 266
DB 303 GAGGDDGGGGLFGAGDG-----GAGGSLTTGAAGNGNAGTSLG-----AAGGAG 352
QY 267 CTGCTGCCACCTGCCCCAGACCGCAAGGTGAGCAAGGTGAGATTTCCAGCAGTGC 326
DB 353 GTGG-----AGGTVGGGKGGAGGAGN-----AGMLFGSG 383
QY 327 ATCGACTACATCAGGACCTTCACTTGGAGCTGAATCTCGAATCCGAACTGGGACCC 386
DB 384 G--GGGTGGTGAAGG-----GGVGSAGMLSGSGSGGAGSGGPGATAGGA----- 431
QY 387 GGGGGCGG-AGGGTGTGCGGTTCGGGCTCCGCTCAGCACCTCAACGGCGAGATCAGCG 445
DB 432 GGAGGAPLIGNGNGNGGSGSGTG-----GVGGAGNAVLINGG 473
QY 446 -CCTACGCGCGAGCGCGATG----- 467
DB 474 EGGIGALAGSGFGFGGILLGADGVNAPESTSPWNLQDILSFINEPTEALTGRPLIG 533
QY 468 ---GTTCTCGGACGATCGCATCTTGTGTGTAAGCGCTCCCGCAGGA-----C 518
DB 534 NGDSGTPTGDDGGAGGWLFGNGG-NGGAGAAGTNGSAG-----GAGGAGGILFG 582
QY 519 CGGCGGACCCAGCCATCCAGGGGCG-----AAGAGGAATTACGTCTCTGTGG 567
DB 583 TGGAGG-----AGVGTAAGAGGAGGAGSAPLIGSGTGGVGAATTTGGVG-----GAGG 633
QY 568 --GTCTCCCCCAAGCGCCCTCG-CCGATCTG-----AGCGAACAACACCGAT--CG 616
DB 634 NAGLLIAGNLGGCCGAGTAVTTTGGAGGTGAGAGLFGANGAGGAGGTGATAGGAGG 693
QY 617 CGCGCCACTCGCCCTTAACTGATCCAGCCTCGGCTGAGGTGAGGCACTGCG-GAGG 675
DB 694 GAGGLYAHGG-----TGGPFGNGSGSTGAGGTGGAGG---PGGLYCAGG 733
QY 676 AGAGGGCGCTCTCTGTCACACTACTAGTACCAGAGACTTTAGGGGTGGGATTCCA 735
DB 734 SGGAGGHH----- 754
QY 736 CTCGTGTGTTTCTATTTTGAAGCAGACATTTTAAATAATGGTCACTGTTGGTGCTT 795

DB 755 -----GSLTLTNASGAGGSGSSLSGKAGA-----GGAG--- 783
QY 796 CTCGATGTTCTGAGAAATTTGCTTTGTTATTTATTAATGATCATCCGACTCAGATA 855
DB 784 ---GSAGLFGSGGAGGNGGYSLNGTGGD-----GTGAGGQITGLRSFGGAGG 830
QY 856 TTGTTTTACAATAG-----TTCTGTGGGCTGTTTTTTTATTATAAATAAATAATAGA 911
DB 831 AGGASDTGAGNGGAGGKAGLYGNGDGGAG-----GDCATSGKGGAGNAVVGNGG 883
QY 912 TGGTCAAAAAA 925
DB 884 NGNAGKAGGTAGA 897
RESULT 9
A70812
Hypochemical glycine-rich protein Rv0833 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 07-Mar-2003
C;Accession: A70812
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Scares, R.; Stulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genom
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70812
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-749 <COL>
A;Cross-references: GB:AL123456; GB:AL123456; NID:93261550; PIDN:CAAL7639.1; PID:92916
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv0833
C;Superfamily: elastin related uncharacterized glycine-rich protein, PE motif containi

Query Match 12.4%; Score 713; DB 2; Length 749;
Best Local Similarity 30.2%; Pred. No. 4.2e-26;
Matches 289; Conservative 21; Mismatches 361; Indels 286; Gaps 41;
QY 23 GTCGCCAAGATCATGAAAGTCCAGTGGCAGCACCGCCACCGCCCGCGGCCCGCCAG 82
DB 3 GNGGAGGSGAPG-AIAGGAGGAG-LIIVGGAGGAG--GDSAVAGVIGAGGAGGALLFG 58
QY 83 CTGC-----GCCCTGAAGCCGCAAGACAGCAGCGGTGCGGCGAGGTGTGCGCTGTC 138
DB 59 AGGAGGSGSGAGGAGGAGGAGGAGLFGAG-GGGFGGFPASTGTGAGGTG-GAGGLF 116
QY 139 TGTCTGACAGAGGTGGCCATCTCGCTGCTCCCG-----GGCGCCGGCGCGCTGCCT 194
DB 117 AS--GGVGGTGGAGSGG-----TGGVGGTGGAGLFGASGAGGAGSGS-GTGGAGG--- 165
QY 195 GCCCTGCTGACAGCA-----GCAGGTAA--ACGTGCTCTACGACATG 239
DB 166 ---TGGAGGLFGAGGAGLGGGQNGHTGGHGGAGSAGLLALGDDGAG---GAGAAATTG 218
QY 240 -AACCGCTGTTTACTCAGCCCTCAAGAGGCTGGTCCCAACCCCTGCCCAGAACCCAG-- 296
DB 219 TGGAGGAGKAGLLFGSGGAGSGGAAAGTFGDTG-----NSGGAGGAGGAKAGLL 267
QY 297 -GTGAGCAAGGTGAGATTCTCCAGCAGCTCATCTACATCAGGACCTTCAGTTGGA 355
DB 268 FSGS---GAGSGGAG-----GFANGSTGGAGG-----AGGGA---GLTNGNGNG 306
QY 356 GCTAACTCGGAATCCGAAGTTGGGACCCCGGGGCGGAGGGGTGCGGCTCCG-GCTC 414
DB 307 GSGGTSVATGAGN--GGAGGAGGAGLIGNGNGG---SGMGDAPGGTGVGIGLL 360
QY 415 GCCTCAGCACCCCT-----CAACGGCGAGA-----TC 440

A;Residues: 1154-1409, 'A', 1411-1420, 'K', 1422-1425, 'R', 1427-1481 <ADA3>

A;Cross-references: EMBL:X06741; NID:93355; PID:929603
 R;Yang, Y.; Adam, R.D.
 Nucleic Acids Res. 22, 2102-2108, 1994
 A;Title: Allele-specific expression of a variant-specific surface protein (VSP) of Giardia
 A;Reference number: S48056; MUID:94301794; PMID:8029018
 A;Accession: S48056
 A;Molecule type: DNA
 A;Residues: 1-56 <YAN>
 A;Cross-references: EMBL:L25059
 A;Experimental source: trophozoites WBA6
 A;Note: the source is designated as Giardia intestinalis
 C;Comment: This translation was produced by PIR staff from information provided by the a
 C;Genetics:
 A;Gene: VSPA6
 C;Keywords: surface antigen; tandem repeat

Query Match 12.6%; Score 725.5; DB 2; Length 1766;
 Best Local Similarity 22.8%; Pred. No. 1.9e-26;
 Matches 370; Conservative 23; Mismatches 461; Indels 767; Gaps 92;

QY 3 GGCCCATCTCTGTTTTCAGC-----CAGTCG-----CCTAGAA-----33
 DB 186 GTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTN 245
 QY 34 -----TCAT-----GAAA-----GTCGCCAGTGG-CAGC-----56
 DB 246 PSDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPY 305
 QY 57 -----ACGCCACCCCG-----CGCGGGGCCCCG-----CTGGCG 89
 DB 306 LKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPCTACAGTADKCT--KCD 363
 QY 90 CTGAA-----GGC-----CGGCA-----AGACAGCGCGGTGGG--120
 DB 364 ANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPC--TACAGTA 421
 QY 121 -GCAGAGTGGTGCCTGTC-----TGCTGA-GCAG-AGCGTG-----156
 DB 422 DKC-----FKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECK 474
 QY 157 CCATCTCGG-----CTCGCGGGC-----GCCGGGC-----184
 DB 475 CNAPTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDA 534
 QY 185 -GCGCTGCTCCCTGCC-----CTGCTG-----GAC--GAGCAGCAG-----215
 DB 535 KECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDD 594
 QY 216 -----GTAACGTGCTGCT-----CTAC-----GACATGACCGCTG 247
 DB 595 SVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAG 654
 QY 248 -----TTACTCAG-----CCTCAAGGAG-----C 267
 DB 655 YTTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDC 714
 QY 268 TGGTGCC-----CACCTGCCCCAGAA-----CGGCAAG-----TGAGC 302
 DB 715 QGSAGYYTDDSVSDAKECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTC 773
 QY 303 --AAGGTGAGATT-----CTCCAGCAGCGTCATC-----329
 DB 774 VSAVDCQGSAGYYTDDSVSDAKECKKNAPCTACAGTA-DKCTKCDANGAAPYLKKTNP 832
 QY 330 ---GACTACATCAG-----GACCTCAGTTGGAGCT-----GAA--361
 DB 833 DPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPCTACAGTAD--KCTKCDANGAAPY 890
 QY 362 -----CTCGAAATCCGAG-----TTGGGACCCCGGGCGGAGGTGCGG-404
 DB 891 LKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPCTACAGTADKCTKCDAN 950

RESULT 8

H70987
 hypothetical glycine-rich protein rv1759c - Mycobacterium tuberculosis (strain H37Rv)
 N/Alternate names: wq22 antigen homolog
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C/Accession: H70987
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
 A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

QY 405 -----GTC-----CGGCTCGCTCAGCA 423
 DB 951 GAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPCTACAGTADKCT 1010
 QY 424 CC-----CTCAACGCCGAGA-----TCAGCCCTCTGACGG 453
 DB 1011 KCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPCT-ACAG 1069
 QY 454 ---C--CGAGC-----GGC--ATCGGTTT-----CTGCGGAC 479
 DB 1070 TADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKNAPC 1129
 QY 480 GATCG-----CATCTTG-----TGTC-----GCTGAAG-----CGC 505
 DB 1130 TACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDAKECK 1189
 QY 506 CTC-CCCAGGAGC-----GGCGGAGCC-----528
 DB 1190 CNAPTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDDSVSDA 1249
 QY 529 --CAGCCATC--CAGGGGGA-AGAGGA-----TTAGTGCTCTGTCG-----567
 DB 1250 KECKKNAPCTACAGTADKCTKCDANGAAPYLKKTNPSPDPTGTCVSAVDCQGSAGYYTDD 1309
 QY 568 -----GTC-----TCCCCAACCGCC-----TCGCCGATCTGAGGAGAAC 605
 DB 1310 SVSDAKECKKNAPKNTAGTCTGFCSDANCERCQDNDVCARCS---TGAPPENGKCP 1365
 QY 606 AGACCGATCGC-----GCCCATCG-----GCCCTTAACCTGAT---641
 DB 1366 AATPCHSSCDGCTENAMTNQADKCTGCKEGRYLKPESAAQSGTCLTAEBCSTDTTHFT 1425
 QY 642 -----CCAGCCTGGGCTGAGGC---TGAG-----GCACCTGG 670
 DB 1426 KEKAGDSKGMCLSCSDATHGITGCKKALKTLGSEASTVWCSECTDKRLTPSGNACLDN 1485
 QY 671 CGAG--GAGAGGGCGCTCTCTGTC-----ACACTAC-----TAGTCAC-CAG 712
 DB 1486 CPAGTYADNINGSVSCASCHATCAECNDANAASCTACYPGYSLLYSGTAGTCVKECTG 1545
 QY 713 A-----GACTTTAGGGGTGGGATCCACTGCTGTTCTATTATTTTGAAGACGAGA 765
 DB 1546 AFITNCADGQCTANV-----GGAKYCAQCKGDGAPIDGIC---TTVAAGRDASV 1592
 QY 766 CATTTTAAAAATGCTACGTTTGGTGC-----TTCTCA--GATTTCTGAGGAAT 814
 DB 1593 C-----TAADGKCT--KCAGEYTLMSGGCGYVAKUPGKSVCTLASNGKCTMCAANNQAPVE 1646
 QY 815 TGC--TTTGTAATTGTATTATAC-----AATGATCACCGACTG-----AGAAATATTGTTT 862
 DB 1647 EKCPECSBCKACNDSNACTECLPGYKAGDKCFKCTASSGNNNQITGVANCVI-----1701
 QY 863 ACAATAGTTCTGTTGGGCTGTTTTTTTGTATTATAACAAATAATTAGATGTTGAAAAA 922
 DB 1702 -CAPPAG-----GNNG-----PVTCTIKTDGNTGGSVKNKSLGTGAI 1738
 QY 923 A 923
 DB 1739 A 1739

```

746.      QY      TCTATTTTGTGA-----AAAGCAGA 765
              :      |||
894      Db      SGGNGGNGNGGDDGNGAHGTGVAQFVPATSLPTPNGGAGGNGGTSGNGGAPGPAGA 953
              :      |||
766      QY      CATTTTAAAA-----AATGTCACCGTTTGGT 791
              |||
954      Db      PGPITGTGNAGSQGIGGGGNGGDKGKGGDGADAVVWFMPTEPQAATCTAGSAGDPTGDN 1013
              |||
792      QY      G-----CTTCTCAGATTCT-----GAGGAAATGCTTT--GTATTGTATATACAAATGAT 840
              :      |||
1014     Db      GPGTPTGSPMVA PPPPTPI TVQQGGDGGAGGTGSTNANDGATGGKGGGE---GGVGS 1069
              :      |||
841      QY      CACCGACTGAGAATATTGTTTACAATAGTCTTG-----TGGGGCTGTTTTTTGTTGTTATT 895
              :      |||
1070     Db      LGGPGGNGTGGNASTGTGTVANAGNGKGGDGGQFGAGGNGGAGGSVTDGSGAGSTAGN 1129
              :      |||
896      QY      BAACAATAAATT-----TAGATCGTCTCAAAAAA 923
              :      |||
1130     Db      GGNGGNATNGTTACGPAGGNGSAGGKGGDGNIAAGA 1166

RESULT 6
T13954
MEGF6 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T13954
R: Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like
A:Reference number: 214126; MUID:98360089; PMID:9693030
A:Accession: T13954
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1574 <NAK>
A:Cross-references: EMBL:AB011532; NID:G3449293; PIN:BA032462.1; PID:G3449294
A:Experimental source: Strain Sprague-Dawley; brain
C:Genetics:
A:Gene: MEGF6

```

Query Match	13.3%	Score 764;	DB 2;	Length 1574;
Best Local Similarity	28.9%	Pred. No. 3.1e-28;		
Matches 302; Conservative	31;	Mismatches 401;	Indels 312;	Gaps 74;

1	QY	GGGGCCCATCTTGTT---TCAGCCAGTC---GGC---AAGAATCATGAAGTC---GC----	46
549	Db	 GWTGIIICNETCPDPTFGKNCSSPC--TCQNGTCDPVLGACRCPPGVSGAHCEDGCPXGF	606
47	QY	-----CAGTGGC-----AGCAC-----C-----GCCACCGC-----	67
607	Db	YKHKRKKCHCANRGRCHRLYGALCLDPGLYGRFCHLACPPWAFPGCESDCLCQSHTR	666
68	QY	---C-----GCCGGGGC---C---CCAGTCGGC---GCTGAAGCCCGCAAGACACGCGAGC	113
667	Db	SCNPXDGSCSKCAFQGBQRCAECSEGFPGGRHRTCPGVACDPVSGECRTCPRGY	726
114	QY	GTTGGGGGGAGGTGGTGGTGTCTCTCTGAGCAGAGCG--TGGCCATCTCGCGCTGCC	171
727	Db	 OGEDCGQEC---PVGTFPGVNCSGSCS-----CVGAPCHRVTG-----ECLCPPGKT---	769
172	QY	GGGGGGC---CGGGGGCGCG---CTCCCTGCCC---TG---CTGGAACGAGCAG--CAGTAAA	220
770	Db	 GEDCAGCPGPRWELGGOEICPACRHHGASCNPETGTCLCLPFGVGRQDTCAGW----	825
221	QY	CGTGC--TGCTTTACAC--ATGAACGGCTGTACTCAGC-----CTCAAGGAG	266
826	Db	 YTGCGQIRCACANDGHCDPTTGRCSAPGWTGLSCQRACDSGHHGWPDCIHPNCVSAGHN	885
267	QY	 CTGTTGCC-----C--AC-----CCTGCCCCAGAAC-----CGCAGGTCGA	300
886	Db	 CDVAVSLCLCBAGVEGPRCEOSCRQGYVGSFCBQKCRCRHGAACDHVSGACTCPAGWFGS	945
301	QY	-----CCAAAGGTGAGATTCTCCACACCGTCATCGACTACATCAGGACCTTCAGTTGGAG	356

```

Db      946 FCEHACPAAGFFGLDSDASCNCAGAPCD--AVTGSC-ICPAGRWPPCAOSCPPLTFGLN 1002
Qy      357 CTGAACTCGGAAATCCGAAGTGTGGACCCCGGGGCC-----GA-GGGTTGCC----GG 405
Db      1003 CS-QICTCFNGASCDSVT-----GOCHCAPGMWPTCLQACPPLYGNKNSHLCSRNGG 1056
Qy      406 TC---CGGGCTC-----CGTC-----AGCACCTCAACGG-----CGAGAT 439
Db      1057 RCDPILGQCITPCBWTGLACENECILPGHYAAGCQLNCSLH-GGI CDRLTGHCLCPAGWT 1115
Qy      440 CAGC-GCCCTGACGGCC-----GA-----GGCGCATG-----CGTTCTGT-CGGA 478
Db      1116 GDKOSSCVSGSTFVHCEEHCACRGASHCHVTGAC-FCPFGMRGPHEQAACPRGWGEA 1174
Qy      479 CGATCGC--ATCTTGTGTCG-TGAGGGC-----CTCCCCCAGGAGA CCGGC 522
Db      1175 CAQRCLCPTNASCHHTVGSRCPPGFTGLSCEAQCPQFTFKDCBEHKQCQPGETWADPA 1234
Qy      523 GGACCACAGCATCCAGGGGGGCAAGAGCAATTACGTGC--TCTGTGGGTC--TCCCCCA 577
Db      1235 SGVCTCAAGYHGTCGL--QRPSG-----RYGPSGEHIKCLNGTCDPATGACYCP 1284
Qy      578 ACGGC-CCTCCCGATCTGAGGGAGACAAGACC--GATCGGGGGCCAC-----TGCGC- 629
Db      1285 AGFIADGSIAUC-----PPGRTPSCAHVCACRGAAACDPVSAGACISPKTGVRC 1336
Qy      630 --CCTTAATGTCACTCAGCTCGGGCTGA-GGCT---GAGCACTGCGAGGAGAGGGCG 683
Db      1337 HGCPQDRFRFGCKELKA--CRNGGLCHATNGSSCSPLGMMGPHCEHA CPAGRYGAACLL 1394
Qy      684 CTC-----CTCTCTGCAC-----ACCTACTAG-----TCACAGA-----GACT 717
Db      1395 CFQNQSCEPTTGACLQGPFYGQAACHSPSPFHGGCORVCECOOQPACPDVPVSGQL 1454
Qy      718 TTAGGGGTGGGATTCACCTGCTGTGTGTTTCTATTATTTTTGAAAAGCACACATT TAAAAA 777
Db      1455 CPAFGHC-----QFCXEGCSSGFGBGDLQQCNCHTVPCDDIPSLGCL--LCPPGR 1502
Qy      778 TGSTCAAGTTTG--GTGCT-TCTCAG 800
Db      1503 TGAACLDLDRGFRFGCALRDGG 1528

RESULT 7
A:1215
A:1215 trophozoite cysteine-rich surface antigen 170 - Giardia lamblia
N:Alternate names: CRP170; cysteine-rich surface antigen CRP170
C:Species: Giardia lamblia
C>Date: 05-Dec-1998 #sequence revision 05-Dec-1998 #text_change 01-Dec-2000
C:Accession: A42125; B42125; S00530; S48056
R:Adam, R.D.; Yang, Y.M.; Nash, T.E.
Mol. Cell. Biol. 12, 1194-1201, 1992
A>Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the
A:Reference number: A42125; MUID:92186850; PMID:1545800
A:Accession: A42125
A:Molecule type: DNA
A:Residues: 1-98 <ADA1>
A:Cross-references: GB:M83937; NID:g159124
A:Experimental source: Trophozoite
A>Note: sequence extracted from NCBI backbone (NCBIN:88421, NCBIP:88427); this C
A:Note: the authors report but do not show 19 tandem repeats of the sequence of
A:Accession: B42125
A:Molecule type: DNA
A:Residues: 1269-1766 <ADA2>
A:Cross-references: GB:M83933; NID:g159122
A>Note: sequence extracted from NCBI backbone (NCBIN:88424, NCBIP:88431); this C
R:Adam, R.D.; Agarwal, A.; Lal, A.A.; de la Cruz, V.F.; McCutchan, T.; Nash, T.
J. Exp. Med. 167, 109-118, 1988
A>Title: Antigenic variation of a cysteine-rich protein in Giardia lamblia.
A:Reference number: S00530; MUID:88089405; PMID:3335828
A:Accession: S00530
A:Molecule type: DNA

```

RESULT 7
A42125 trophozoob
N;Alterna
C;Species
C;Date: C
C;Access
C;Access
Mol. Cell
A;Title:
A;Referen
A;Access
A;Molecu
A;Residu
A;Cross-r
A;Experim
A;Note: s
A;Note: s
A;Access
A;Molecu
A;Residu
A;Cross-r
A;Note: s
A;Cross-r
Mol. Cell
J. Exp. M
A;Title:
A;Referen
A;Access
A;Molecu

222 GNGDGGGGAAGAVIGTSGATGAGDGGHGTGAAGNGGTGGAGGGGIDGVGGGTGG----- 277
 161 CTGCGCTGCCG-----GGCGCGGGGGCGCGCTCCCTGCTGTGACGAGCAGCAG 215
 278 -TGNGNGNGAIGAGDAGGSGNSGNG-GIG-----GKGNAGAGGAAGSNG 323
 216 GTAAACGTGTCTCTACGACATGAACGGCTGTTACTACGCTCAAGAGCTGTGTGCC 275
 324 GTVGAAGTGGDGGGGAAGAAATAGSNGGAGTG-----SAGNGGTGGG----- 367
 276 ACCTGCGCCAGAACCCAGAGGTGAGCAAGTGG----- 310
 368 ---GSGAGAGDGTGGVGGGNGAGDEVGAGAGGSGPNTSPGNGGGGGGAG 424
 311 GATTCTCCAGACGTATCATGACTATCATCAGGAGACTT-----CAGTTGAGCTGACTCG 365
 425 GAAGAGAGGAGGANGTAGNGGGAGGTGGAGAASSATNGSGGAGGTGGDGGSGGAGTG 484
 366 GNATCCAGATTGGGACCCCGGGGGCCGAGG-GCTGCGGTCCGGCTCCGCTCAGCAC 424
 485 GAGGTGAAGDGGGGGGAGAGAGGGGGAGGAGGTGGNGTGGTAGTAGNAGNGGA- 543
 425 CCTCAAGGGGA-GATCAGCGCCCTGACGCCCGAGCGGCATCGTTCTCTGCGGACGATC 483
 544 ---AGKGGAGGCGGTGGTG--GQGGAGGDDGAGGTGGRTVGGGTVPAGSGGGGAG 597
 484 GCAT-----CTGTGTCTG-----CTGAAGCGCTCCCGAGGACC 519
 598 GGGAGGGGAGDGGSGGGDAGTGGNGNGNGNNGTGGAGN-----GGGAN 648
 520 GCGGAGCCCAAGCCATCAGAGGGGCAAGAGAAATTA--CGTGTCTGTGGTCTCCCCCA 577
 649 GGAG-----AGSGGGTGGNGGAGDAGDAGNGNGTNG----- 686
 578 ACGCGCTCGCGGATCTGAGGAGAAACAGACCGATCGCGGCCACTGCGCCCTTAAT 637
 687 -----GNGSGNGIAGMGNGGAGGTGSG--NGNGSGGNGGAGMGNGSGTSGD 734
 638 GCATCCAGCC---TGG-GGCTGAGGTGAGGCACCTGGCG----- 672
 735 GGAGNGGAATGTTGGDGGTGTGGTGGSGGTGGDGGNGGNGADTANMTAAGDGGN 794
 673 -----AGGAGGGCGCTCTCTCTGCACACCTACTAGTACACAGACTTTAGGGGT 726
 795 GGDGFGGAGAGGG-GLTAGANGTGGGQ-----GAGDGGNGATGHHPLTDPGNGGT 849
 727 GGGATT-----CCACTGCTGTGTTCTATTTTT-----AAAAGCACACATTT 770
 850 GNGGTGTGTGAGTGLGGTGGDGGNGGNGGTGGGEGVGGAGTGGAGNGDGGTGG 909
 771 TAAAAAATGGTCACTT-----TGGTGCTTCTCAGATTCTTCTGAGGAATGCTTT 820
 910 TGGDGGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 964
 821 GTATTGTATTACATATCATCAGCTAGTGAAG--TATTGTTTACATAGTTCTGTGG- 877
 965 GNGGTGGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1023
 878 ---GGCTGTTTTTTTATTAAACAAATATTTAGATGCTGAAAAA 921
 1024 GGTGGGGNGGTGNGGKGTGSGGAGDGGKAPAGGTGGAGDGGGA 1070

RESULT 5
 E70806
 hypothetical glycine-rich protein RV3507 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: E70806
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Rellwell, T.; Gierles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: E70806
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-1381 <COL>
 A: Cross-references: GB:AL020222; GB:AL123456; NID: g3261554; PIDN: CAA17744.1; PID: g2924.
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: RV3507
 C: Superfamily: collagen alpha 1(IV) chain

Query Match 13.3%; Score 764.5; DB 2; Length 1381;
 Best Local Similarity 31.3%; Pred. No. 2.7e-28;
 Matches 350; Conservative 19; Mismatches 463; Indels 285; Gaps 44;

QY 1 GGGGCCCATCTCTTTTTCAGCCAGTCCCAAGATCATGAAAGTCG-----CCAGTGGCA 54
 141 GGGG---APGAAGSGGNGSAGLWNGGAGAGGAGGAGGAGGNGGNGGWLFGAGGTGGIG 197
 QY 55 GCACCGCCACCGCCCGCGC---GGGCCCGAGCTGCGCTGAAGGCGCGCA---AGAC 106
 198 GTGAPGAMGTGGNGNGALLIGGGGLGGAGCMGTG-GGTGGTGGNGGALLIGAGGV 256
 QY 107 AGCGA-GCGGTGCGGGGAGGTGTGCTGT-CTGTCTGAGCAGAGCGTGGCCATCTCG 164
 257 GAGGIGGCGTGAAGAGAGGTGGG-CAGGLFVNGDGGCAGGQ-GDGAAGDAASAGG 314
 QY 165 CGCTCCCG-GGCGCCGGG---CGCGCTCCTGCTGACGAGCAGCAGGTTAA 220
 315 TGKGGGGGDTGGAGAGPVLFGHGGG-----MGQGGTGGNGAGGDTTVA 367
 QY 221 CTGTGCTGTACGACATGACCGCTGTTACTACGCTCAAGAGCTGGTGGCCACCC- 279
 368 AGTGGGCTGGAAGA---GGAAGAGALTSGLAGGVAGGTGTGTGTGGAGADAATAVVG 424
 QY 280 -----TGCCCCAGNACCGCA--GCTGAGCAAGTGGAGATTCTC----- 317
 425 FGANGDPFAGKGGNGGIGGAATVGVAGDGTGGKGTGGAGAGANDAGTGNPGKG 484
 QY 318 -----CAGCAGCTCATCGACTACATCAGG-----GACCTTCAGTTGGA 355
 485 GGGGIGGAGGAGAGTNGGHAGNTGDDGDTGGNGGTGGVNGADNTLNPDTPGA 544
 QY 356 GCTGAATCGAATCGAAGTTGGACCCCGGGGCGAGGGCTGCGGCTCCGGCTCC 415
 545 GEPGAGGAGGAGGAGGPGGTGG-----TGNGNGNGNGNG-----GGNGNGG- 589
 QY 416 GCTCAGACCTCAACGCGCAGATCAGCGCCCTGACGCGCGAGGCGCATG-----CGTT 470
 590 -----NGNAGNNSTNAPVG--GEGGAGDGGAGGAGGAANGTAGSQGTG 633
 QY 471 CTGCGGACGATCGCATCTTGTGCTGTAAGCGCTCCCCAGGAGCCGCGGACCCCA 530
 634 GVGGGGAGGNGGKAGTGNNGFVDEAG-----FSGAGNGGAGGANG 684
 QY 531 GCATCCAGGGGCAAGAGAGATTACGTCTCTGT--GGGTCTCCCCCAACGCGCTCGC 588
 685 GTGGGGGDDG-AGGTGAG-----NGIPTGTETPAGT-----GAKG-----GD 726
 QY 589 CGCATCTGAGGAGAACAGACCG-ATCGCGGCCACTGCGCCCTTAATCTGAT-CCAGC 646
 727 GDDGAGGAGNAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 778
 QY 647 CTGGGCTGAGCTGAGGACTGGCGAGG-----GAGGCGCTCTCTCTGCACA 697
 779 SGDDGGKGGGSGGTGGGAGPAGGAGTGGSGHAGGAGGAGTGAAGTTTTFNGNGN 838
 QY 698 CTTACTAGTCAACAGAGA-----CTTTAGGGGTGGGATTCACATCTGTGTT-- 745
 839 AGDGGNGN---AGAGGNGSGDFFGNTTSGASGSGGNGGAGTAGSGAG-GTGGTGL 893


```

Db 424 AGSSGAGGTNGSGAGGCGGAGCA-----CGAGADNPTGTGTGDTGDTGGAAGAGG 476
QY 367 AATCCGAGTTG-----CGACCCCGGGCGGAGGGCT-----GCCGGTCCGGG 411
Db 477 AGAAGTGGTTCGMITGNAGVGA-----GGGGGGGAGAGADADQFGATGTTGTFAGG 531
QY 412 CTCGGCTCAGCACCCCTCAACGGCGAGATCAGCCCTGACCGCGGAGGCGG--CATCGGTT 470
Db 532 AGGAG-----GAGSSGAGGTNGSGAGGTGGGAGGAGGAGADNPTGIGGT 579
QY 471 CCT-GCGGAGCA-TCCGATCTTGT-CTCCGCTGAAGCGCCTCCCGCAGGACCGGGGACC 527
Db 580 GBDGTTGGAAGAGAGAGAAAGTGTGGMITGNAGV-----GAGGGGGGAGAGAGA 632
QY 528 CCAGCCATCAGGGGGCAAGAGGAAT---TACGTGCTCTGTGGTCTCCCGCAACCGCC 584
Db 633 DADQFAT---GGTGFAGGAGGAGGAGGSSAGGTNSGAGGTGRQSGTGGAGGAGAD 688
QY 585 TCGCCGATCTGAGGAGAAACAGACCGAT-CGGCGGCACTCGCGCCCTTAATCGATCC 643
Db 689 NPTGIGG---TGGDGGTGGAGAGGAGGAAGTGTGGMIGTTG---NAGVG-----733
QY 644 AGCTGGCGGTGAGGCTGAGGCACTGGCGAGGAGAGGCGCTCTCTCTGCACACCTACT 703
Db 734 -----GAGSSGAGGTNGSGAG--GTDGGGAGGAGGAD---NPTG-----I 773
QY 704 AGTCACAGACACTTTAGGGGGTGGGATTCACCTCGTGTCTTCTATTTTTGAAGACA 763
Db 774 GGT-----GGDGGTGGAGAGGAGGAA-----GTGCTGGMIGTTGNAGVGAGGQ 818
QY 764 GACATTTTAAAA---ATGCTACGTTTGTGCTTCTCAGATTCTTGAGGAAATTCGTT 819
Db 819 GDDGAGGAGADADQFGATGGTGFAGGAGGAGGSSGSSCAGGT---NGSGAGGTGQVV 875
QY 820 TGTATTGTATTAATCAATGATCACCAGTACGAGAAATTTGTTTACAAATAGTTCT-----873
Db 876 AGGA--GISFSNGSNGGTGGTGGVGG--TGGDGNAGTGADEFKGTGGTGGTGGSGGA 931
QY 874 -----GTGGGGCTG-----TTTTTTTCTTATTAACAATAATTTAGATG 913
Db 932 GSGGANFNGTGTGTGTGGKGLNTDGLSSATSGTGTGTGTGGTGGGDDSGAGTG 991
QY 914 GTGAAAAAAA 924
Db 992 GTGGAGGNAGA 1002

```

RESULT 2

```

fibroin - Chinese oak silkmoth
C:Species: Antheraea pernyi (Chinese oak silkmoth)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31328
R:Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, Antheraea pernyi
A:Reference number: Z20995
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2639 <SEZ>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C:Genetics:
A:introns: 14/3

```

```

Query Match 15.0%; Score 863; DB 2; Length 2639;
Best Local Similarity 25.9%; Pred. No. 1.4e-32;
Matches 316; Conservative 41; Mismatches 545; Indels 318; Gaps 31;

QY .1 GGGGGCCATCTGTTTCCAGCATCG-----CCAGAAATCATGAAGTCCCGAGTG 51
Db 1128 GSAIAAAAAAAAAAGAGGGGGYGGWGDGGYGSIAIAAAAAAAAAAGAGGGGGY 1187

```

```

QY 52 G-----CAGCACCGCCACCGCGCGCGCGG-----CCCCAGCTGCGCGCT 91
Db 1188 GYGSIAIAAAAAAAAAAGAGGGGGYGGWGDGGYGSIAIAAAAAAAAAAGAGGAGS 1247
QY 92 GAAGGCGG-----GCAAGACAGCGAGCGGTGCGGCGCAGGTGGTGGCTGTCTCTCT 143
Db 1248 GGYGGYGGYGSIAIAAAAAAAAAAGAGGAGGAGGGYGGWGDGGYGSIAIAAAAAAAAA 1307
QY 144 GAGCAGAG-----CTGGCCATCTCGCGCTGCGG-----CGTGGCCATCTCGCGCTGCGG 173
Db 1308 GSGAGGCGDGGYGSIAIAAAAAAAAAARRAGHRAAGSIAIAAAAAAAAAAGAGGAGS 1367
QY 174 GCGCCCGGGCGCGCTGCTGCGCGAG-----CAGCAGGTAAACGT 223
Db 1368 GGYGGWGDGGYGSIAIAAAAAAAAAAGAGGAGGAGGGYGGYGSIAIAAAAAAAAAAGS 1427
QY 224 GCTGC-----TCTACGACATGAACGGCTG-----TTACTCACCGCT 259
Db 1428 GAGYGGGYGGWGDGGYGSIAIAAAAAAAAAAGAGGAGGAGGGYGGYGSIAIAAAAAA 1487
QY 260 CAAGGAGCTGTCGCCACCTGCCCCAGAACCGCAAGTGA--GCAAGGTGGAGATTCTC 317
Db 1488 AAAGAGAGGYGGYGSIAIAAAAAAAAAAGAGGAGGAGGGYGGWGDGGYGSIAIAAAA 1547
QY 318 CAGCACGTCA-----TCGACTACATCAGGACCTTCAGTTGGAGCTGAATCGAATCC 371
Db 1548 AAAAAAGAGGAGGAGGGYGSIAIAAAAAAAAAARRAGHRAAGSIAIAAAAAAAAAA 1607
QY 372 GAAGTTGG-----GACCCCGGGCGGAGGCTGCGCGTCCCGGCTCC 415
Db 1608 GAGGAGGGYGGWGDGGYGSIAIAAAAAAAAAAGAGGAGGAGG-----YWGDDGY 1658
QY 416 GCTCAGCACCTCAACGCGGAGATCAGCG-----CCTCAGCGCCCGAGCGG 462
Db 1659 GSDSIAIAAAAAAAAAAGS--GAGGAGGGYGGWGDGGYGSIAIAAAAAAAAAAGSGAGRGD 1717
QY 463 CATCGTTTCT-----GCGACCATCGCATCTTGTCTG-----496
Db 1718 GGYGSIAIAAAAAAAAAARRAGHRAAGSIAIAAAAAAAAAAGAGGAGGGYGGWDDGY 1777
QY 497 ---CTGAAGCGCTCCCGCAGGACCGCGG-----ACCCGACCATCCAGGGGG--543
Db 1778 GSDSIAIAAAAAAAAAAGSAGGGYGGYGGYGSIAIAAAAAAAAAAGAGGAGG 1837
QY 544 -----CAAGAGAAATTAGTCTCTGTG-----GTCTCCCCA-----577
Db 1838 VGGYGGWGDGGYGSIAIAAAAAAAAAAGSGAGRGDGGYGSIAIAAAAAAAAAARRA 1897
QY 578 ---ACGCGCTCGCGCATCTGAGGAG-----AACAAAGACCGATCG 616
Db 1898 GHDRAGSIAIAAAAAAAAAAGAGGAGGGYGGWGDGGYGSIAIAAAAAAAAAAGSGAG 1957
QY 617 GCGGCCACTGGCGCCTTAACCTGATCCAGCTGGGGCTGAGGCTG-----A 662
Db 1958 GAGGGYGGWDDGYGSIAIA-----AAAAAAGAGGAGGAGGGYGGWGDGGYGSIAIAAAA 2013
QY 663 GGCATCTGGCGAGGAGAGG--GGCTCTCTCTGACAC-----CTACTAGTCAACAGAGAC 716
Db 2014 AAAAAAGAGGAGRGDGGYGSIAIAAAAAAAAAARRAGHRAAGSIAIAAAAAAAAAAAA 2073
QY 717 TTTAGGGGTGGATTCACCTGCTGTGTTTCTATTGTTTGAAGACAGAC-----766
Db 2074 ASGAGGGYGGY-----GYGSIAIAAAAAAAAAAGSGAGGAGGYGGYGGY 2123
QY 767 ---ATTTAAAAAATGTCAGT-----TTGCTCTCTCAGATTTCTGAGGAATTCG 817
Db 2124 SDSAIAIAAAAAAAAAAGSGAGGGYGGWGDGGYGSIAIAAAAAAAAAAGSGAGRGD 2183
QY 818 TTTGATTGTTATATTACA-----ATGATCACCGACTGAGAAATTG-----858
Db 2184 GYGSIAIAAAAAAAAAARRAGHRAAGSIAIAAAAAAAAAAGSGAGSGYGGYGGY 2243

```


Qy 913 G-----GTGAAAAAAAA 925
Db 1519 GCAGGCCATGTGCACGAGACA 1539

Search completed: May 7, 2004, 15:25:49
Job time : 73 secs

219 AACGTGCTGTCTACGACATGAACGGCTGTACT-----CACGCTCAAGG-- 264
462 CGCGCCCTGGCT-----CATGACGCGGCTCCCGCTGGCGCGCGCCCGCGGCT 517
265 -----AGCTGGTCC-----CACCTGCCC-CAGAAACGCAAGGTGACAAAGG 306
518 GTGAATGCGACTCGCCCTCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGG--ACGTGG 574
307 TGGAGATTCTCCAGCAGCTCATCA-----CTACA-----TCAGGG-- 342
575 TAGGGGATGCCAGCTCACTGCAATGGCAGTTGGCGGCTCTCCAGTTCCCTCTGGTC 634
343 ACCTTCAGT-----TGGAGCTGAAT-----CG-GAATCC-----GAAGTTGGACCC 384
635 ACCTGCTGCTGATGGTGGCTGTGTGAGTCCGAGCATCCGCTGGAGAAGCTGG-----C 690
385 CCGGGGCGAGGGCTGCGCGTCCGG-----CTCCGCTCAGCA---CCCTC 428
691 CCAAGCAC--AGAGCAGCGGCGCAGAGAGCGTGAGCAGCACTCGGGACGCGCCGG 749
429 AACCG-----CGAGATC-AGCGCCC--TGACGGCCGA-GCGGCGATGCGTTC-----CT-- 473
750 GCGGGTGAACGAGCTCGCGCGCGCGCGCGGAGGACGAGGCGCGCA-GCGCGCGGACTGG 808
474 -----GCGGAC-----GATCGC--ATCTTGTCTC-----GCTGAGC--GCC 506
809 AAGAGCAAGACGCGCGCTGCGCGCGCTGAGCGCTGGAGCAAGCTGAAGCAGGCGC 868
507 ----TCCCCCAGG-----ACC--GGC--GGACC-----CCAGCCATCCAGGGG- 542
869 TGGGTCTCCAGCGCGGGGCGCCAGGCGGGGATCTGCAGGTCCGCGCCCGCGGGAC 928
543 -----GCAAGAGAAATTACG-----TGCTCTG-----TGGGTCTC-CC 574
929 ACCCGCAGGCGGAAGCCCTGSCCGAGCCCGCCAGGACGCGATTGCGCCGGAACTCGGG 988
575 CCAACGC-----GCCTC-----GCCGA--TCTGAGG-----GAG 602
989 CCCAGCCCGAGCCACCCGAGGAGTACGTGTACCCGACTACCGTGGCAAGGGCTGGTG 1048
603 AACAAAG-----AC-CGATCG-----CGG-----CCACTGC--GC 629
1049 GACGAGAGCGGCTTGTGTACGATCGGAGAGAGTTTCGCGCGGCGCCCTCGGCTGC 1108
630 CCTTAAC--TGATC-----CAGCCTGGGG--CTGAGGCTGAG- 663
1109 CCGTCCCTGTGACGAGAGGGGCGCTGTGCGCGCAGCCCGAGTGCCTCGAGGCTGCAC 1168
664 --GCACT-----GGCAGGAGAG--GGCGCT--CCTCTCTGC-----AC 696
1169 CCGCGCTCATCCAGTCCGACGAGCCAGTGTCTGCCCGCAGTGCATGAAGAGAGAAC 1228
697 ACCTACTAGTACCAG-----AGACTTTAGGG-----GTGG-GATTCCTACT 737
1229 TACTCGAGT--TCCGGGCAAGACCTATCAGACTTTGAGGAGTTCTGTGTCTCATG 1287
738 CGT---GTGTTTCTATTT-----TTTGAAG--CAGACATTTTAA 774
1288 CGAGAGGTGTGCTGTGAAGCCAAACGTTAGGTGTATGACAGTGTGAG-CGTGTCCC 1346
775 AAA-----TGCTAC-----GTTTGGTGT-TCTCAGATTTCTGAGGA 811
1347 AGACGAGGTGTGAGCCCTGTGTACGAGCCTGTATCAGTGTCTCCCA-----TCTGCAA 1402
812 AAT-----TGCTTTGTATTGTATTACATGATC---ACCGACTGAGAA-TATTG 858
1403 AATGTTCCAACTGCTTTGCA--GAAACCGCGGTGATCCCTGTGGCAGAGAGTGAAG 1459
859 TTTTCAATAGTTC--TGTGGGCGTGTGTTTGTGTTTATTAACAAAT----AATTAGAT 912
1460 ACTGACGAGTGCCACCATATGCCACTGTACTATG-AGGAGGCACATGGAGATCGAGCG 1518


```

Db 402 AACCCGGGCGCGCGCGCGTGTACCCCTGGCGCGCTGCGAGCCGCGGTCCGGCC 461
Qy 219 AACGTGCTGTACGACATGAACGCGTGTACT-----CAGCCTCAAGG-- 264
Db 462 CGCGCCCTGCGCT-----CATGACGCGCGTCCCGCTGCGCGCGCGCGCGCGCT 517
Qy 265 -----AGCTGCTGCC-----CACCTGCCC-CAGAACCGCAAGGTGAGCAAGG 306
Db 518 GTGAATGCGACTCGCCCTCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGG--ACGTGG 574
Qy 307 TGGAGATTCTCCAGCACGTCATCGA-----CTACA-----TCAGGG-- 342
Db 575 TAGGGATGCCAGCTCCACTGCGATGCGAGTTGGCGCGCTCTCCAGTTCCCTCTGTC 634
Qy 343 ACCTTCAGT-----TGGAGCTGAAT--CG-GAATCC-----GAAGTTGGACCC 384
Db 635 ACCTGCTGCTGATGTTGGTCTGTGTCAGTCCAGATCCCGTGGAGAGCTGG----C 690
Qy 385 CCGGGGCGCGAGGCTGCGCGTCCGG-----CTCGGTCAGCA---CCCTC 428
Db 691 CCAGGCACC-AGAGCAGCGCGCGCGAGAGAGCGTGAGCAGCGCACTCGGGAGCGCCGG 749
Qy 429 AACGG-----CGAGATC-ACGCGCC--TGACGCGCGA-GSCGCGATGCGTTC---CT-- 473
Db 750 GCGCGGTGAACGAGCTCGCGCGCGCGCGAGAGCGTGAGCAGCGCACTCGGGAGCGCCGG 808
Qy 474 -----GCGGAC-----GATGCG--ATCTTTGTGTC-----GCTGAAGC--GCC 506
Db 809 AAGAGCAAGAGCGCGCGTGGCGTGGCGCGTGGAGCGTGGAGCAAGCTGAAGCAGCGCC 868
Qy 507 -----TCCCCCAGG-----ACC--CGC--GGACC-----CCAGCATCCAGGG- 542
Db 869 TGGGTCTCCAGGCGCGCGCGCGCAAGCGCGGGATCTGAGTTCGCGCGCGCGCGGAC 928
Qy 543 -----GCAAGAGGAATTACG-----TGCTCTG-----TGGGTCTC-CC 574
Db 929 ACCCGCAGCGAGCGCTGGCGCAGCGCGCGCGAGCGAGTGGCGCGGAACTCGCG 988
Qy 575 CCAACGC-----GCTC-----GCCGA--TCTGAGS-----GAG 602
Db 989 CCAACGCCGAGCGCGCGAGGAGTACGTGTACCGCGACTACCGTGGCAAGGCGTGGT 1048
Qy 603 AACGAG-----AC-CCATCGG-----CGG--CCACTGC--GC 629
Db 1049 GACGAGCGCGCTCGTGTACCGATCGGGAGAGTTGCGCGCGCGCGCGCTCGGCTGC 1108
Qy 630 CTTTAAAC--TGATC-----CAGCTGGG--CTGAGCTGAG- 663
Db 1109 CCGTGCCTGTGCACCGAGGAGCGCGCTGTGCGCGCAGCGCGAGTCCCGAGGCTGCAC 1168
Qy 664 --GCACT-----GGCGAGGAG--GCGCT--CCTCTCTG-----AC 696
Db 1169 CCGCGCTGCATCCAGCTGCACAGCGCAGTGTGCGCGCGAGTCCCGAGGAGAGAAC 1228
Qy 697 ACCTACTAGTACCAG-----AGACTTTAGGG-----GTGG-GATTCACCT 737
Db 1229 TACTGCGAGT--TCGGGCAAGACCTATCAGACTTTGAGGAGTTCGTGTCTCCATG 1287
Qy 738 CGT---GTGTTCTATT-----TTTGAAG--CAGACATTTTAA 774
Db 1288 CGAGAGGTGCTGCTGTGAAGCCACCGGTGAGGTGCTATGACAGTGTGAG--CGTGTCCC 1346
Qy 775 AAA-----TGCTCAC-----GTTTGTGCT-TCTCAGATTCTGAGGA 811
Db 1347 AGACGAGGTGTGAGCCCTGTGTACGAGCGCTGATGCTGTGCTCCCA---TCTGCAA 1402
Qy 812 AAT-----TGCTTTGTTATTATTATTATGATC---ACGACTGAGAA-TATTG 858
Db 1403 AATGTTCCAAATGCTTTGCA---GAAACCGCGGTGATCCCTGCTGCGAGAGTGAAG 1459
Qy 859 TTTTACATAGTTC--TGTTGGGCTGTTTTTTTGTATTATTAACAAT---AATTTAGAT 912

```

```

Db 1460 ACTGACGAGTGCACCATATGCTACTGTACTTATG-AGGAAGGCACATGAGAAATCGAGCG 1518
Qy 913 G-----GTGAAAAAAA 925
Db 1519 GCAGGCCATGTGCACGAGACA 1539

RESULT 15
US-10-123-155-515
; Sequence 515, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 515
; LENGTH: 1942
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-515

Query Match 47.6%; Score 2744.5; DB 14; Length 1942;
Best Local Similarity 41.1%; Pred. NO. 1.5e-129;
Matches 625; Conservative 0; Mismatches 27; Indels 621; Gaps 96;

Qy 3 GSCC-----CAT-TCTGTTTCAGCC-AGTCCCAAGAAT--CATGAAGTC--- 44
Db 42 GSCCGCGTCCGGGTTCTTTTCTCCGACGCGCCACGCTGCCAGACATTCCGCGC 101
Qy 45 -GCCAG--TGGC-AGCACCGCCACCGCGCGCGG- 74
Db 102 TGCCGGGTCTGAGAGCTCCCGAACCCCTCCGCGAGAGAGCGAGCGCGCAGGTT 161
Qy 75 GSCCCGCTGCGG- 92
Db 162 GSCCCCGGGGCGCTTGTGTCGGAAGCGGGGACGAGCCGAGGAGTGCAGCTG 221
Qy 93 AAGG-----CCGGCA--AGACAGC----- 109
Db 222 AGGCGAGCGCGGGCACTGACGCGAGTTGGGGCCGCGACTACCGGCGAGCTGACAGCGCAT 281
Qy 110 GAGCG-----GTGCGGCGAGGTGTCGCG--TGTCGT 140
Db 282 GAGCGACTCCCGAGAGAGCGCCCTAGCCCGTGTGCGCGCAGCGCGAGCGAGTGGG 341
Qy 141 TGTGAGCAGAGGTGG-CCATCTCGCG-----TGCGG-----GGGCG----- 177
Db 342 GCTGGGCTGTAGTGTCCCGCCACGCGGGTCCGCGCGCCCGCCAGGATGGCGCTGCG 401
Qy 178 ----CGGG--GCGCGCGCTG-----CCTGC--CCTGCTG---GACGAGCAGAGGTA 218
Db 402 AACCCGGGCGCGCGCGCGCGTGTGTACCCCTGCGCGCGCGCTGCGAGCCGCGCGTCCGGCC 461

```


Db 2020 GCGGTGGCGGTGGCGCGGGGTTCCCGGAGAGGCGCTGCTGGCGCCCTGATGACGAAC 2079
QY 211 AGCAGGTAA---CGTGTCTGCT---CTAGCATGAAC---GGCTGTATA 250
Db 2080 GGTGGGCAAGGCGACGCTGCTGAGGGCGGGCCCGACGACTGGACTCGGGGTGCTG 2139
QY 251 CTCAGCCT---CA-----AGGAGCTGTGCTCCACCTGCTCCCGACGAACCGC 293
Db 2140 CCGACCGCGAGCAGACGCGCTGCGCGAGAGCGCTGCCCACTCCGACCGCGACGCC 2199
QY 294 AAGG-----TGAGCAAGGTG---GAGATTCTCCAG---CACGTCACTGACTACA 336
Db 2200 CAGCCCTGGGCGCCCGCTGGGACACGAGCCACCCCTGCTCCGGGCTCCGCTTCA 2259
QY 337 TCAG-GGACCTTCAGTTGAGCTGAACCTCGGAATCGAAG-----TTGG--- 379
Db 2260 TCCTCCTCTGCTGCTGGCGC-CGCGCGGGCGCCCGAGAGCCCGCGCTGGGA 2318
QY 380 ---GACCCCGGGGGCG---AGGCTG-CGGTCCGGG----- 412
Db 2319 GCGGACCGCGAGCGCGCTGCTATGCTCCGCGCGCGCGCGCTCCCGAGCTT 2378
QY 413 TCCGCTCAGCACCT-----CAAGCGCGAG---ATCAGCGCCCTGACGG-- 453
Db 2379 CCGCTCACCCCGACCGCGAGCGCGCGCGGTGGTGTCCGGGCC--ACGGCG 2436
QY 454 -CGAGS---CGGATGCTGCTGCGGAGAGTGCATCTTGTGCTGAAGCGCTCC 509
Db 2437 CCCTTGACCGACCT-----CAGCGCGGATGGCTCCCGCGCGCTGGAGC-CGCGC 2489
QY 510 CCC--AGGG--ACC-----GGC-----GACCCCA-----GCCATCCAG-G 540
Db 2490 CCGGAGCGGCTGAGGAGGCACTGGGCGCCCGACGCGCTCCCGCGCGCACCTGG 2549
QY 541 GGCA-----AGAGAAATTAGTCTCTGTGGTCTCCCGCAAGC-GCCT 585
Db 2550 CCGCACCCACACGTTCAAAGCGCGAGCGCGCGCTGGGACCGCGACCGCGGTGCA 2609
QY 586 CGCC-----GGATCGA-----GGAGAACAGACCGATCGG 617
Db 2610 CGCCCGCGCGGACAGACCTTGCGCCACCTCTCCCTATGGGGGC--GACAGACTG 2668
QY 618 CGGCCACTGCGCCTTAAC-----TGCA-----CCAGCTGGGG--- 652
Db 2669 CGCCCGCGTGCCTAGGCGCGGGCGCCCGGATGCTTGGCAGTCCGACCGACGGAAC 2728
QY 653 -----CTGAG-GCTGAGGCACTG----- 670
Db 2729 CAGAGCGAGACGGTGCAGAACCGCGGGCGCGGGGCACTCGAGTGGGTGCTCAA 2788
QY 671 -----CGAGGAGAGGCGCTCCTCTC-----TGAC-ACC 699
Db 2789 GTCCCCCGGACCGACCGCGAGTGGGGGCGCCCTCGGCAACAGGAGCAAC 2848
QY 700 TACTAGTC-----ACGAGACT-----TTAGGGGTG----- 727
Db 2849 AGCTCGCCTCCCGCTACCGCGGCGCGAGACGCTGAGACGGTTTGGGGGTGGTGGC 2908
QY 728 -----GGATT-----CC-----ACTCGTGTGTT---CTATTTT 755
Db 2909 GGGAGGACTTGTATGATTTGAGTTGACCTTATGCGGTAGGTTTGGTTTTTTT 2968
QY 756 GAA-----AAGCAGACATTTTAAATAATGGTCA-----CGTT-----TGGT 791
Db 2969 GCAGTTTTGTTTCTTTTGGGGTTTCTAACCAATTGCACAACCTCGTTCTCGGGTGGC 3028
QY 792 -----GCTT---CTCAGAT-----TTCTGAGGAATTG- 816
Db 3029 GGCAGGAGGGGAGCTTGACCGCGGTGGGAATGGGGGCCACAGCTGCAGACCTAG 3088
QY 817 -----CTTGTATTGTATAT-----TACAAATGATCAACGAC-----TGAGATAT- 856

Db 3089 CCCTCCCCCAACCCCTGGAAGAGTCCCTCCCAACCCAGGCCCCCTGGCGTGTGGGTGTG 3148
QY 857 -----TGTTT---TACAAAGTCTGTG---GGGCTGTTTTTTTGT 892
Db 3149 CGTGGGTGCGTGGCGCTGTTCTGTGCAAGGGCGCGGAGGTGGGC-GTGTGTGTGCG 3207
QY 893 ATTAAACAATAATTATAGATGGTGAAAAA 926
Db 3208 TGCCAGGAAGGCTGCTGTGGCGTGTGTGTCAA 3241

RESULT 13

US-10-140-864-543
; Sequence 543, Application US/10140864
; Publication No. US20030207419A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Deanoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C184
; CURRENT APPLICATION NUMBER: US/10/140,864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-864-543

Query Match 48.4%; Score 2790.5; DB 15; Length 3721;

Best Local Similarity 39.7%; Pred. No. 1.3e-131;

Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;

QY 1 GGGGCC-----CATTC-----GTTTC-----AGCAG--TCGCCAAG 31
Db 1723 GGGGCCAGCACCTCAGGCTTAGGGGACTGCACAGGACTCCTGGGGCCAGCCTCTCCGAG 1782
QY 32 AATCATG-----AAAGTCGCCAGTGGCAGC----- 56
Db 1783 GACCGCGGGGCTGGTGTGGTGAACCTGCTGTAACGTGCGTGGCGCCTTCGTG 1842
QY 57 -----ACCG-----CC-ACCGCCG-----CGGGGCC--CC--AGCTGC-- 86
Db 1843 GTGGAGCGGTGTGTCGGGCTTCAGCGTGGGTGTTCTGGGCTCCGTAAGCGCGG 1902
QY 87 GCGGTGAAGCCGCG-----AGACA-----GCGAGCGGTGCGGGCGAGTGG 129
Db 1903 GAGCTG--GCCCGCGGAGAGCAAGAGGCGCATCTGTCGCGCAGCGGCGGGCGAGCGG 1960
QY 130 TGC---GCTGTCT---GTCTGAGC-AGAGCGTGGC-CA---TCTC----- 163
Db 1961 TGTGTAGC-GTCAGCGCTGGGCGAGCGAGGAGGAGGCGCATCTGTCGCGCAGCGGCGGGCGGGA 2019
QY 164 -GCGGTGCGCG-GGCGCGGGCGCGCTGCTGCTGCTGCTGCTG-----GACGAGC 210
Db 2020 GCGGTGCGGTGGCGCGGGGTTCCCGGAGGCGCTGCTGCGGCCCTGATGCAGAAC 2079

QY 817 -----CTTTGATTGTATAT-----TACAATGATCACCGAC---TGAGATAT- 856
Db 3089 CCTTCCCCCAAGGTCCTCCCAACCCAGGCCCTGCGGTGTG 3148
QY 857 -----TGTTT-----TACAATAGTTCGTG---GGCTGTGTTTGTGTT 892
Db 3149 CGTGGGTGCGTCCCGCTGTTGCGTGCAGAGGCGCGAGGTGGGC-CTGTGTGTGCG 3207
QY 893 ATTAACAATAATTAGATGTTGAAAAAARA 926
Db 3208 TGCCAGGAGGCTGCTGTGGCGGTGTGTCAA 3241

RESULT 12

US-10-140-805-543
; Sequence 543, Application US/10140805
; Publication No. US20030207417A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvatroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tamas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P333081C176
CURRENT APPLICATION NUMBER: US/10/140,805
CURRENT FILING DATE: 2002-05-07
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 543
LENGTH: 3721
TYPE: DNA
ORGANISM: Homo Sapien

US-10-140-805-543

Query Match 48.4%; Score 2790.5; DB 15; Length 3721;
Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;

QY 1 GGGGCC-----CATTCT-----GTTTC-----AGCCAG--TCGCCAAG 31
Db 1723 GGGGCCAGCACCTCAGGCTTAGGGAGTGCACAGGACTCTCGCGGCCAGCCTCTCCGAG 1782
QY 32 AATCATG-----AAAGTCGCCAGTGGCAGC----- 56
Db 1783 GACCGCGCGGGTGTGTGCTGAACCTGTGTAACTGTCTGTCTGTGTGTGTGTGTGTGTGT 1842
QY 57 -----ACCG-----CC-ACCGCGC-----CGCGGCC-CC---AGCTGC-- 86
Db 1843 GTGGAGCCGT 1902
QY 87 GCGCTGAAGCCCGGC-----AAGACA-----GCGAGCGGTGTGTGTGTGTGTGTGTGTGT 129
Db 1903 GAGCTG--GCCCGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1960
QY 130 TGC---GCTGTCT---GTCTGAGC-AGAGCTGTGC-CA---TCTC----- 163
Db 1961 TGCTGAGC-GTCAGCGCCTGT 2019
QY 164 -GGCTGTGCGG-GGCGCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 210

QY 164 -GGCTGTGCGG-GGCGCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 210
Db 2020 GCGGT 2079
QY 211 AGCAGTAA-----CGTGTCTCT-----CTAGCAGTGAAC-----GGCTGTGTTA 250
Db 2080 GGT 2139
QY 251 CTACAGCT---CA-----AGSAGCTGT 293
Db 2140 CCACGCGCGAGCAGAGCGCGCTGCGCGAAGAGGCTGCCACTCCGACCGCGCACCC 2199
QY 294 AAGS-----TGAGCAGGTG-----GAGATTTCTCAG---CAGCTCATGACTACA 336
Db 2200 CACGCCCTGGGCGCGCGCTGT 2259
QY 337 TCAG-GGACCTCAGTTGAGTGAAGTCTGGAATTCGGAAG-----TTGG----- 379
Db 2260 TCCT 2318
QY 380 ---GACCCCGGGGGCG---AGGGTGTG---CCGGTCTCGGGC----- 412
Db 2319 GCGACCGCGAGCGCGCTCTATGT 2378
QY 413 TCCTCTCAGCACCT-----CAAGCGCGAG---ATCAGCGCCTGACGG-- 453
Db 2379 CCGCTTACCCCGCAGCGCAGCG 2436
QY 454 -CGGAGS---CGGATGCTGT 509
Db 2437 CCTTGTGACCGAGCT-----CAGCGCGCGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2489
QY 510 CCC--AGGG--ACC-----GGACCCCA-----GCCATCCAG-G 540
Db 2490 CCGACCGGCGAGCTGAGGAGCGCACGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2549
QY 541 GGGCA-----AGAGAAATTAAGT 585
Db 2550 CCGACCCACAGCTTCAACAGCG 2609
QY 586 CGCC-----GGATCTGA-----GGAGAACAGACCGATCGG 617
Db 2610 GCGCGCGCGCGCGAGACTTGTGCGCCACT 2668
QY 618 CGGCCACTTGGCGCTTAAC-----TGCAAT-----CGACCTGGGG-- 652
Db 2659 CGCCCGCGCTGCGCTTGT 2728
QY 653 -----CTGAG-GCTGAGGCACTGG----- 670
Db 2729 CAGGAGCGAGAGACGCTGCCAGAACCG 2788
QY 671 -----CGAGGAGAGGGCGCTCTCTC-----TGAC-ACC 699
Db 2789 GTCCCGCGCGACCCACCGCGGAGTGGGGCG 2848
QY 700 TACTAGTC-----ACGAGAGCT-----TTAGGGGTG----- 727
Db 2849 AGCTGCGCT 2908
QY 728 -----GGATT-----CC-----ACTCGTGTGTTT---CTATTTT 755
Db 2909 GGGAGGACTTGTCTATGATTTGAGTTGACCTTATGCGCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2968
QY 756 GAA-----AAGCAGACATTTTAAATAATGTCA-----CGTT-----TGST 791
Db 2969 GCAGTTTGTGTTTCTTTTGTGCGTTTCTTAAACAATTGACAACTCCGTTCTCGGGGTGCG 3028
QY 792 -----GCTT---CTCAGAT-----TTCTGAGGAAATG- 816
Db 3029 GGCAGSAGGAGGCTTGAAGCGCGGTGGGAAATGGGGGCGCACAGCTGCAGACCTTAAG 3088

QY	130	TGC---GCTGTCT---GTCTGAGC-AGAGCGTGGC-CA---TCTC-----	163
Db	1961	TGCTGAGC-GTCAGCCGCTTGGCGAGCGCAGGCGCAGGGTCCGGGGCCCGGGCGGA	2019
QY	164	-GGCTCTCCGG-GGCGCGGGGGCGGCTTCCCTGCCCTGCTG-----GACGAGC	210
Db	2020	GGCGGTGGCGGTGGCGCGGGGTTCCTCCCGGAGGCCCTGCTGGCGCCCTGATG	2079
QY	211	AGCAGGTAA----CGTGTCTCT-----CTACGACATGAAC-----GGCTGTTA	250
Db	2080	GGCTGGGCCAAGGCCACGCTGCTGCAGGGCGGGCCCCACGACGTGACTCGGGGCT	2139
QY	251	CTCAGGCT--CA-----AGGAGTGTGTGCCACCTGTGCCAGAACCCG	293
Db	2140	CCACGCCCGCAGACGCGCTGCCGAGAAAGGCTGCCCACTCCGACACCGCACCC	2199
QY	294	AAG-----TGAGCAGGTG-----GAGATTCTCAG---CAGTTCATCGACTACA	336
Db	2200	CAGCCCTGGGCCCCCGCGCTGGGACACGCGCCACCCCTGTCTCCGGCTCCGCTTCA	2259
QY	337	TCAG-GGACCTTCAGTTGGAGCTGAATCTCGGAATCCGAAG-----TTGG---	379
Db	2260	TCCTCCCTCTCTGCTGTGGCGCC-CGCCCGGGCCCCGAGCAGCCCCCGGCTGGGA	2318
QY	380	---GACCCCGGGGGCG---AGGGTG-CCGGTCGGGC-----	412
Db	2319	GCCGACCCCGACGGCGCCTCTATGTGCTCCCGCGCGCGCGCTCCACGGGACCTT	2378
QY	413	TCGCTCAGCACCCCT-----CAACGGCGAG---ATCAGCGCCTTGAACG--	453
Db	2379	CCGCTCACCCCCACGACGCGGACGCGGGGTGTGTCTCCGGGCC--ACGGGC	2436
QY	454	-CGAGG---CGGATCGTTCTGCGGACATCGCATCTGTGTCTGTGAAGCGCTCC	509
Db	2437	CCCTTGACCCAGCCT-----CAGCGCCGATGGGCTCCCGGGCGCTTGAGC-CG	2489
QY	510	CCC--AGGG--ACC-----GGCACCCCA-----GCCATCCAG-G	540
Db	2490	CCCGACGGGACGCTGAGGAGGCCACTGGGCCCCACGCCCTCCGGCGCCACCTGG	2549
QY	541	GGGCA-----AGAGGAATTACGTCTCTGTGGTCTCCCCCAAGC-GCCT	595
Db	2550	CCGACCCACACGTTCAACAGCGCGGAGGCCCGGCTGGGAGACCGCCCGCGCTGCCA	2609
QY	586	CGCC-----GGATCTGA-----GGGAGAACAGACCGATCGG	617
Db	2610	CGCCGGCGGGCACAGACTTGGCCCCACCTCTCCCTATGGGGGGC-GACAGAGCTG	2668
QY	618	CGGCCATGCGCCCTTAAC-----TGCA-----CCAGCCTGGG---	652
Db	2669	CGCCCCGCTGCCCTAGCGCGGGGGCCCCCGATGSCCTTGGCAGTGCACGCCACGGAAC	2728
QY	653	-----CTGAG-GCTGAGGCACTGG-----	670
Db	2729	CAGGAGCGAGAGACGGTGCCAGAACGCGGGGCCCGGGGCAATCCGAGTGGGTGCTCAA	2788
QY	671	-----CGAGAGAGGGCGCTCTCTC-----TGCAC-ACC	699
Db	2789	GTCCCCCGCGACCCACCGCGAGTGGGGGGCCCCCTCCGCCACAAGGAAGCACACCC	2848
QY	700	TACTAGTC-----ACCAGAGACT-----TTAGGGGGTG-----	727
Db	2849	AGCTCGCCCTCCCCCTACCCGGGGCCGAGGACGCTGAGACGGTTTGGGGGTGGGTGGC	2908
QY	728	-----GGATT-----CC---ACTCGTGTGTTT---CTATTTTTT	755
Db	2909	GGGAGGACTTTGCTATGAGTTGAGTTGACCTTATGCGCGTAGTTTGGTTTTTTTTT	2968
QY	756	GAA-----AAGCAGACATTTTAAAAAATGGTCA-----CGTT-----TGGT	791
Db	2969	GCAGTTTTTGGTTTCTTTTGGGTTTTTCTAACCAATTGCACAACTCCGTTCTCGGGGTGGC	3028

```

QY 792 -----GCTT-----CTCAGAT-----TTCTGAGGAAATTG- 816
Db 3029 GGCAGGAGGGGAGGCTTTGGACGCCCGTGGGGAATGGGGGGCCACAGCTGCGAGACCTTAAG 3088
QY 817 -----CTTTGTATTGTATAT-----TACAATGATCAACCGAC---TGAGAATAT- 856
Db 3089 CCCTCCCCCACCCTCGAAAGGTCCCTCCCCAAACCCAGGCCCTGGCGTGTGTGGGTGTG 3148
QY 857 -----TGTTT---TACAATGTTCTGTG---GGGCGTGTTTTGTGT 892
Db 3149 CGTGGGTGTCGTGCGCTGTTCGTGTGCAAGGGGCCGGGAGGTGGGC-GTGTGTGTGCG 3207
QY 893 ATTAACAACAATAATTTAGATGTTGAAAAAATA 926
Db 3208 TGCAGCGAAGGCTGCTGTGGGCGTGTGTCAA 3241

RESULT 10
US-10-141-756-543
; Sequence 543, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C200
; CURRENT APPLICATION NUMBER: US/10/141,756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-756-543

```

Query Match	48.4%	Score	2790.5	DB 15	Length	3721			
Best Local Similarity	39.7%	Pred. No.	1.3e-131						
Matches	609	Conservative	0	Mismatches	302	Indels	623	Gaps	83
QY	1	GGGGCC	-----CATCT	-----GTTTC	-----AGCCAG	-----TCGCCAAG	31		
Db	1723	GGGGCCAGACCTCAGCCTTAGGGGACTGCACAGGACTCTCGGGGCCACCCCTCTCCGAG	1782						
QY	32	AATCATG	-----AAAGTCGCAGTGGCAGC	-----	56				
Db	1783	GACCGCGGGGCTGTGTGCGTGAACCTCTGTTAACTGCTCGTGTGGGGGCTTCTGTT	1842						
QY	57	-----ACCG	-----CC-ACCGCCGC	-----CGCGGCC	-----CC	-----AGCTGC	86		
Db	1843	GTGGAGCCGGTGTTCGGCTTCAGCGTGGGCTGTTCTGTGGCCCTCCGTAGCGCCG	1902						
QY	87	CGCTGAAGGCCGC	-----AAGACA	-----CGAGCGGTGCGGGCGAGGTGG	129				
Db	1903	GAGCTG	-----GCCCGCGCAAGACAAAGAGGCCATCTCTGGCGGACCGGGCGGGCGAGCGG	1960					
QY	130	TGC	---GCTGTCT	---GTCTGAGC	---AGAGCGGTGGC	---CA	---TCTC	-----	163

2969	Db	GCAGTTTGGTTCTCTTTTGGGGTTTTCTAACCAATTGCACAACTCCGTTCTCGGGGTGGC	3028
792	Qy	-----GCTT-----CTCAGAT-----TTCTGAGGAATTTG- 816	
3029	Db	GGCAGCAGGGAGGCTTGGACGCCGTTGGGAAATGGGGGCCACAGCTCAGACCTTAAG	3088
817	Qy	-----CTTTGTATTGTATAT---TACAATGATCACCAGC---TGAGAAATAT- 856	
3089	Db	CCCTCCCCCAACCCCTGGAAAGGTCCTCCCAACCCAGGCCCTCGCGGTGTGGGGTGTG	3148
857	Qy	-----TGTTT---TACAATAGTTTGTG---GGGCTCTTTTTTTTGTTT 892	
3149	Db	CGTGCCTGTGCGGCCGTGTTCTGTGTCAAGGGCCGGGAGGTGGC-CTGTGTGTGCG	3207
893	Qy	ATTAAACAAATAATTAGATGGTGAATAAAAAA 926	
3208	Db	TGCCAGCGCAAGGTGCTGTGGCGGTGTGTGTCAA 3241	

g. Scores

1843	Db	GTGGAGCCGTGTGTCTCCGGCTTCAGCGTGGGCTGTCTCGTGGGCTCTCGTGAAGCGCGG	1902
87	Qy	GCCTGAAGCCCGC--AAGACA-----GCGACGCTGCGGGCGAGCTGG	129
1903	Db	GAGCTG--GCCCGGCAAGGAAGAGAGCCATCTGGGCGACGGGGCGGGGAGAGCGG	1960
130	Qy	TGC--GCTGTCT--GTCTGAGC-AGAGCGTGGC-CA--TCTC-----	163
1961	Db	TGCTGAGC-GTCAGCGCCTGGGCGAGCGAGGGCGCAGGGTCTCCGGGGCCCGGGCGGA	2019
164	Qy	GCCTGCGCG-GGCGCGGGGCGCGCTGCTGCGCTCGCTG--GAGAGC	210
2020	Db	GCGGTGGCGGTGGCGCGGGTTCCCGGAGGCGCTGTGGGGCCCCGTGATGCAAGAC	2079
211	Qy	AGCAGGTAAA---CGTGTCTCT-----CTACGACATGAAC---GGCTGTGA	250
2080	Db	GGCTGGGCAAGGCCACGCTGTCAGGGCGGGCCCCACGACCTGGACTCGGGCTGCTG	2139
251	Qy	CTCAGCCT--CA-----AGAGCTGTGCCCCACCTGCGCCCCCAGAACCGC	293
2140	Db	CCACGCCCGAGCAGACGCCGCTGCCGAGAGCGCTGCCACTCCGCAACCCGACCCC	2199
294	Qy	AAGG-----TGAGCAAGGTG---GAGATTCTCCAG--CACGTATCGACTACA	336
2200	Db	CAGCCCTGGGCCCCCGCGCTGGGACCAAGGCCACCCCTGTCTCCGGCTCCGCTTCA	2259
337	Qy	TCAG-GGACCTTCAGTTGAGCTGAACTGCGAATCCGAAG-----TTGG---	379
2260	Db	TCTCTCCTCTGCTGCTGGCGCC-CGCGCGGCGCCCGAGCAGCCCCCGCGCTGGGA	2318
380	Qy	--GACCCCGGGGCGCG--AGGGCTG-CCGCTCGGGG-----	412
2319	Db	GCGACCCCGACGCGCGCTCTATGTGCCCGCGCGCGCGCTCCACGGCGACTT	2378
413	Qy	TCCGTCAGCACCCCT-----CAAAGGAG--ATCAGCGCCCTGAAG--	453
2379	Db	CCCGTCACCCCACGCGAGCGCGGACCGCGCGGGTGTGTCTCGCGGCC--ACGGC	2436
454	Qy	-CCGAGG--CGCATGGTTCCTGCGAGCATCGCATTTGTCTGCTGAAGCGCTCC	509
2437	Db	CCCTTGAACCAAGCT-----CAGCCCGATGGCTCCCGCGCCCTGGAG-CCGC	2489
510	Qy	CCC--AGGG--ACC-----GGACCCCA-----GCCATCCAG-G	540
2490	Db	CCCGACGGGACGCTGAGGAGCCACTGGGCGCCACGCGCCCTCGGCGCCACCTGCG	2549
541	Qy	GGCA-----AGAGGAATTAAGTCTGTGGTCTCCCCCAAGC-GCCT	585
2550	Db	CGCACCCACAGTTCAACAGCGGAGGCCCGCTGGGACCGCCACCGCGCTGCCA	2609
586	Qy	CGCC-----GGATCTGA-----GGGAGAACAGACCGATCGG	617
2610	Db	CGCCCGCGGGCAGAGCTTGGCCCACTCTCCCTCTATGGGGGGC-GGACAGGACTG	2668
618	Qy	CGGCCACTGGCGCTTAAC-----TGCAT-----CCAGCTGGG---	652
2669	Db	CGCCCCCGTGCCTTAGCGCGGGGCCCCCGATGGCTTGGCATGTCAGCCACGGGAAC	2728
653	Qy	-----CTGAG-GCTGAGGCACTGG-----	670
2729	Db	CAGGAGCGAGACCGTGCAGAACCGCGGGGCGCGGGCAATCCCGATGGGTGCTCAA	2788
671	Qy	-----CGAGGAGAGGGCGTCTCTC-----TGAC-ACC	699
2789	Db	GTCCCCCGCGACCCACCGCGGAGTGGGGGGCCCCCTCGCGCACAGGAAGCACACC	2848
700	Qy	TACTAGTC-----ACCAGAGACT-----TTAGGGGGTG-----	727
2849	Db	AGCTGCGCTCCCGCTAACCGGGGCGCAGACGCTGAGACGGTTTGGGGGTGGGTGGG	2908
728	Qy	-----GGATT-----CC-----ACTCGTGTGTTT---CTATTTTTT	755

RESULT 7

```

RESULI /
US-10-158-790-543
; Sequence 543, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; PENDING FILING DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-158-790-543

```

	Query Match	48.4%	Score 2790.5	DB 14	Length 3721
	Best Local Similarity	39.7%	Pred. No. 1.3e-131		
	Matches 609	Conservative 0	Mismatches 302	Indels 623	Gaps 83
QY	1	GGGGCC-----CATTC-----GTTC-----AGCAG--TCGCCAAG	31		
DB	1723	GGGGCCAGACCTCAGGCTTAGGGAGTCGCAGAGCTCTCGCGGCGAGCCTCTCCGAG	1782		
QY	32	AATCATG-----AAAGTCGCCAGTGGCAGC-----	56		
DB	1793	GACCGCGGGGCTGGTCTCGGTGACCTGCTGTGTAACGTCGTGGTGGCGGCTTCGTG	1842		
QY	57	-----ACCG-----CC-ACCGCGG-----CGGGGCC-CC---ACGTCG--	86		
DB	1843	GTGGAGCCGTGGTCTCCGGTCTCAGCGTGGGCTGGTTCGTGGGCCCTCGTAGCAGCGCG	1902		


```

QY 32 AATCATG-----AAAGTCGACGAGTGGGCG----- 56
Db 1783 GACCGCGCGGGGTGGTGTGCTGAAACCTGCTGGTAAAGTCTGGTGGCGGCTTCTGTG 1842
QY 57 -----ACCG-----CC-ACCGCGCG-----CGCGCGCG-CC-----AGCTGC-- 86
Db 1843 GTGGGAGCGTGGTGTCCGGCTTCAGCGTGGGCTGGTGGTGGGCTTCGGTGGGCGG 1902
QY 87 GCGCTGAAGCGCGCG-----AAGACA-----GCGAGCGGTGGCGGAGGTGG 129
Db 1903 GAGCTG--GCGCGCGCAAGGACGAGGCGCCATCTGCGCGACGCGGGCGGCGAGGCGG 1960
QY 130 TGC---GCTGTCT---GTCTGAGC--AGAGCTGGC--CA---TCTC----- 163
Db 1961 TGTGTAGC--GTAGCGCGCTGGCGAGCGAGGCGCAGGCTCCGGGGCGCGCGGCGGA 2019
QY 164 -GCGCTGCGCG--GGCGCGCGGGCGCGCTGCTGCTGCTGCTGCTG-----GAGGAGC 210
Db 2020 GCGGCTGGCGGTGGCGCGGGGTTCGCCCGGAGGCGCTGCTGCGCGCCCTGTATGAGAAC 2079
QY 211 AGCAGGTAA--CGTGTCTCT-----CTACGACATGAAC-----GGCTGTATA 250
Db 2080 GGTGTGGCCAGGACGCTGCTGTGAGGGGGGCGCCACGACTGGACTCGGGGCTGTGTG 2139
QY 251 CTCACGCT---CA-----AGGAGCTGGTGGCCACCTGCTGCTGCTGCTGCTGCTG 293
Db 2140 CCGACGCGCGAGCAGCGCGCTGCGCGCAAGAGCGCTGCCACTCGCAGCCCGCACCC 2199
QY 294 AAGS-----TGAGCAAGGTG---GAGATTCTCAG--CAGGTCATCGACTACA 336
Db 2200 CAGCGCTGGCGCCCGCGCTGGGACCGCGCCACCCCTGTCTCCGCGCTTCGCTTCA 2259
QY 337 TCAG--GGACCTTCAGTTGAGCTGAACCTCGGAATCGAAG-----TTGG---- 379
Db 2260 TCCTCCCTCTGTGCTGTGGCGCC--CGCGCGGGCGCCCGAGCAGCCCGCGCGCTGGGA 2318
QY 380 ---GACCCCGGGGGCGG---AGGGTG--CGGTCGCGGCG----- 412
Db 2319 GCGGACCGCGGCGCGCTCTATGCTGCCCGCGCGCGCGCGCTCCACGCGACTT 2378
QY 413 TCCGCTCAGCACCTT-----CAACGCGCGAG-----ATCAGCGCGCTGACGG-- 453
Db 2379 CCGCTTACCCCGCACCGCGCGCGCGCGCGCGCGGCTGCTGCGCGCGCG--ACGGCG 2436
QY 454 -CGAGG---CGGATCGCTGTGGGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 509
Db 2437 CCCTTGGACCGAGCT-----CAGCGCGCGGATGGCTCCCGCGCGCGCTGGAGC--CGCG 2489
QY 510 CCC--AGGG--ACC-----GGC-----GGACCCCA-----GCCATCCAG-G 540
Db 2490 CCGAGCGGCGAGCTGAGGAGCGCACTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2549
QY 541 GGGCA-----AGAGGAAATTAGTGTCTGTGGGTCTCCCGCGCGCGCGCGCGCG 585
Db 2550 CCGCACCCACAGTTCAACAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2609
QY 586 CGCC-----GATCTGA-----GGGAGAACAGACCGATCGG 617
Db 2610 CGCGCGCGCGGCGACAGACTTGGCGCCACCTCTCCCTATGGGGGCG--GGAAGAGCTG 2668
QY 618 CGGCGACTGCGCCCTTAC-----TGCAT-----CCAGCTGGGG--- 652
Db 2669 CGCCCGCGGTGCTCTAGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAC 2728
QY 653 -----CTGAG--GCTGAGGCACTGG----- 670
Db 2729 CAGGAGCGAGAGAGCGGTCCAGACCGCGGGGCGCGGGCGCGCGCGCGCGCGCGCG 2788
QY 671 -----CGAGGAGAGGCGCGCTCTCTC-----TGCAC--ACC 699
Db 2789 GTCCCCCGCGGACCCACCGCGGAGTGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2848

```

```

QY 700 TACTAGTC-----ACCAGAGACT-----TTAGGGGGTG----- 727
Db 2849 AGCTCGCCCTCCCTTACCCGGGCGCAGGACGCTGAGACGCTTTGGGGTGGTGGCG 2908
QY 728 -----GGATT-----CC-----ACTCGTGTGTT---CTATTTTT 755
Db 2909 GCGAGGACTTTGTATGATTTGAGTTGACCTTATGCGGTAGGTTTGGTTTTTTTT 2968
QY 756 GAA-----AAGCAGACATTTTAAAAAATGGTCA-----CGTT-----TGGT 791
Db 2969 GCAGTTTGGTTCTTTTGGGTTTTCTTACCAATTGCACACTCCGTTCTCGGGTGGC 3028
QY 792 -----GCTT---CTCAGAT-----TCTCAGGAAATG-- 816
Db 3029 GCGAGCAGGCGGAGGCTTGGACCGCGTGGGGAATGGGGGCGCACAGCTGCAGACCTAAG 3088
QY 817 -----CTTTGTATTGTATAT-----TACAATGATCACCAGAC---TGAGAATAT- 856
Db 3089 CCCTCCCCCAGCTGGAAGGTCCTCCCAACCCAGGCGCTGCGGTGTGGTGTG 3148
QY 857 -----TGTTT---TACAATAGTTCTGTG-----GGGCTGTTTTTTTTTGT 892
Db 3149 CGTGGTGTGGTGGCGCTGTTCTGTTGTCGAGGGGCGCGGAGGTGGGCG--GTGTGTGGC 3207
QY 893 ATTAACAAATAATTTAGATGTTGAAAAAATAA 926
Db 3208 TGCCAGCGAAGGCTGCTGTGGGCGTGTGTCTCA 3241

RESULT 4
US-10-140-472-543
; Sequence 543, Application US/10140472
; Publication No. US2003013888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zenin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-140-472-543

Query Match 48.4%; Score 2790.5; DB 14; Length 3721;
Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;

QY 1 GGGGCG-----CATTTCT-----GTTTC-----AGCCAG--TCGCCAAG 31
Db 1723 GGGGCGAGCACCTCAGGCTTAGGGAGCTGCACAGGACTCTCGCGGCGCAGGCTCTCCGAG 1782
QY 32 AATCATG-----AAAGTCGACGAGTGGGCG----- 56

```

Db	2789	GTCCCCCGGAGCCACCCGCGGGAGTGTGGGGGGCCCCCTCCGCCACCAAGGAGACACAACC	2848
QY	700	TACTAGTC-----ACGAGAGACT-----TTAGGGGGTG-----	727
Db	2849	AGCTCGCCCTCCCTACCCGCGGCGCAGACGCTGAGACGGTTTGGGGGTGGGTGGGC	2908
QY	728	-----GGATT-----CC-----ACTCGTGTGTTT-----CTATTTTTTT	755
Db	2909	GGGAGGACTTGTATGGAATTCAGGTTGACCTTATGCGGTAGGTTTGTGTTTTTTTTT	2968
QY	756	GAA-----AAGCAGACATTTTAAAAAATGTGCA-----CGTT-----TGGT	791
Db	2969	GCAGTTTGGTTTCTTTTGGGTTTTTCTTAACCAATTCACAACTCCGTTCTCGGGGTGC	3028
QY	792	-----GCTT-----CTCAGAT-----TTCTGAGGAATTG-	816
Db	3029	GGCAGCGAGGGAGGCTTGACCGCCGTTGGGGAATGGGGGCCACAGTGCAGACCTAAG	3088
QY	817	-----CTTTGTATTGTATAT-----TACAATGATCACCGAC-----TGAGAAATAT-	856
Db	3089	CCCTCCCCACCCCTGGAAAGTCCCTCCCAACCCAGGCCCTGTGGGTGTG	3148
QY	857	-----TGTTT-----TACAATAGTTCTGTG-----GGGCTGTTTTTTTGT	892
Db	3149	CGTGGTGTGGTGGCGGTGTTTGGTGTGCAAGGGGCCGGGAGGTGGGC-GTGTGTGTGGC	3207
QY	893	ATTAAACAATAATTTAGATGTGTAAAAAATAA 926	
Db	3208	TGCCAGCGAAGGCTGCTGTGGCGCTGTGTGTCAA 3241	

RESULT 3
 US-10-146-731-543
 ; Sequence 543, Application US/10146731
 ; Publication No. US20030129692A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: ACIDS ENCODING THE SAME
 ; CURRENT APPLICATION NUMBER: US/10/146,731
 ; CURRENT FILING DATE: 2002-05-15
 ; Prior Application removed - See file Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 543
 ; LENGTH: 3721
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-146-731-543

Query Match	48.4%	Score 2790.5	DB 14	Length 3721
Best Local Similarity	39.7%	Pred. No. 1.3e-131		
Matches 609	Conservative 0	Mismatches 302	Indels 623	Gaps 83

QY 1 GGGGCC-----CATCT-----GTTTC-----AGCCAG--TCGCCAAG 31
 Db 1723 GGGGCCACAGACCTCAGGCTTAGGGGAGTCACAGAGACTCTCGGGGCCAGCCTCTCCGAG 1782

1 GGGGCC-----CATTC-----GTTTC-----AGCCAG--TCGCCAAG 31
1723 GGGCCAGCACCTCAGGCTTAGGGAGCTGCACAGGACTCCTCGCGGCAGACTCTCCGAG 1782
32 AATCATG-----AAAGTCGCCAGTGGCAGC-----56
1783 GACCGCGGGGGTGCTGTGCTGTGAACCTGTGGTAACTGCTCGGTGCGCGCTTCGTG 1842
57 -----ACCG-----CC-ACCGCGC-----CGCGGGCC-CC---AGTGC--86
1843 GTGGAGCCGTGTGTTCGGCTTCAGCGTGGCTGTGTTGTTGGGCTCCGTGAGCGGCG 1902
87 GCCTGAAGCCCGC---AAGACA-----CCGAGCGGTGGCGGCGAGGTGG 129
1903 GAGCTG--GCCCGCGCAAGGACAAAGAGGCCATCTGTGCGCACCGGCGCGCGAGGCGG 1960
130 TGC---GCTGTCT---GTCTGAGC-AGAGCGTGGC-CA---TCTC-----163
1961 TGTGAGC--GTACGCCGCTGTGGCGAGCGCAGGGCGCAGGTCCTCGGGGCGGGCGGA 2019
164 -GGCTGCCGG--GGCGCGGGGCGCGCTGCTGCCCTGCTG-----GACGAGC 210
2020 GCGGTGGCGGTGGCGCGGGGTTCCTCCGAGGCGCTGCTGGCGCCCTGATGACAGAAC 2079
211 AGCAGGTAAA---CGTGCTGCT-----CTACGACATGAAC---GGCTGTTA 250
2080 GGTGGGCAAGCCACGCTGCTGCAGGGCGGGCCCAACGACTGACTCGGGGTGCTG 2139
251 CTACGGCT--CA-----AGGACTGTGCCACCTGTCCCGCCAGACCGC 293
2140 CCCACGCCGAGCAGACGCGCTGCCGCAAGAGCGCTGCCACTCGCACCGCACCC 2199
294 AAGS-----TGAGCAAGTG---GAGATTCTCCAG--CACGTCACTCGACTACA 336
2200 CACGCCCTGGCCCCCGCGCTGGGACACCGCCACCCCTGTCTCCCGCCCTCGCTTCA 2259
337 TCAG-GGACCTTCAGTTGAGCTGAATCGAATTCGAG-----TTGG---379
2260 TCTTCTCTCTGCTGCTGGCGCC--CGCCCGGGCCCCCGAGACGCCCGCGCTGGGA 2318
380 ---GACCCCGGGGCG--AGGGCTG--CCGTCGGGC-----412
2319 GCCACCCCCGACGCGCGCTTATGTGTCGGCGCGCGCGCTCCACCGGCGACTT 2378
413 TCGCTCAGCACTT-----CAGCGCGAG-----ATCAGCGCCCTGACGG---453
2379 CCGCTCACCCCCACGACCGCGGACCGCGCGGGGTGTGTCGCGGCC--ACGGGC 2436
454 -CGGAG--CGGATGCGTTCCTGCGGACGATCGCATCTGTGTGCTGAAGCGCTCC 509
2437 CCTTGACCCAGCT-----CAGCGCGCATGGCTCCCGCGGCGCTGGAC--CGCC 2489
510 CCC--AGGG--ACC-----GGC-----GACACCCA-----GCCATCAG-G 540
2490 CCGACGGGCAGCTTGAGGAGGCGCACTGGGCGCCCCACGCGCCCTCGGCGCCACCTGG 2549
541 GGGCA-----AGAGGAATTAGCTGTCTGTGGGTCTCCCCCAAGC-GCCT 585
2550 CCGACCCACAGTTTCAACACGCGGAGGCCCGGCTTGGGACCGCCACCGCGCTGCCA 2609
586 CGCC-----GGATCTGA-----GGGAGAACAGACCGATCGG 617
2610 CGCCCGCGGGGCAAGACTTGGGCCACTCTCTCCCTATGGGGGGC--GACAGGACTG 2668
618 CGGCCACTCGCCCTTAAC-----TGCAAT-----CCAGCTTGGG---652
2669 CGCCCCCGTGCCTTAGCGCGGGGGCCCCCGATGCTTGGCAGTGCACGCCACGGGAAC 2728
653 -----CTGAG-GCTGAGGCACTGG-----670
2729 CAGGAGCAGAGACGCTGCAGAACCGCGGGGCGCCGGGCACTCCGAGTGGGTCTCAA 2788

QY	671	-----CGAGGAGGGCGCTCCTCTC-----TGACAC-ACC	699
DB	2789	GTCCCCCGCGACCCACCCGCGGAGTGGGGGGCCCCCTCGCCACAAAGACACAACC	2848
QY	700	TACTAGTC-----ACCAGAGACT-----TTAGGGGGTG-----	727
DB	2849	AGCTGCCCTCCCTACTACCGGGGCGCAGGACGCTGAGACGGTTTGGGGGTGGGTGGGC	2908
QY	728	-----GGATT-----CC-----ACTCGTGTCTT-----CTATTTTTT	755
DB	2909	GGGAGGACTTGTCTATGGAATTGAGGTTGACCTTATGCGGTAGGTTTTGGTTTTTTTTT	2968
QY	756	GAA-----AAGCAGACATTTTAAAAAATGTCA-----CGTT-----TGGT	791
DB	2969	GCAGTTTTGGTTCTTTTTCGGGTTTCTAAACCAATTGCACAACTCCGTTCTCGGGGTGGC	3028
QY	792	-----GGTT-----CTCAGAT-----TTCTGAGGAAATTG-	816
DB	3029	GGCAGCAGGGAGGCTTGGACGCCGGTGGGAATGGGGGCCACAGTGCAGACCTAAG	3088
QY	817	-----CTTTGTATTCTATAT-----TACAATGATCACCAC-----TGAGAATAT-	856
DB	3089	CCCTCCCCCACCCTGGAAAGTCCCTCCCAACCCAGCCCTCGCGTGTGGGTGTG	3148
QY	857	-----TGTTT-----TACAATAGTTCGTG-----GGGCTGTTTTTTTGT	892
DB	3149	CGTGGTGTGCGTGGCCGTGTTCTGTGTGCAAGGGGCGGGAGGTGGGC-GTGTGTGTGCG	3207
QY	893	ATTAACCAATAATTTTAGATGTTGAAAAAATAA	926
DB	3208	TGCCAGGAGGCTGCTGTGGGCGTGTGTCTAA	3241

RESULT 2

US-10-123-155-543

Sequence 543, Application US/10123155

Publication No. US20030068794A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

APPLICANT: DeForge, Laura

APPLICANT: Desnoyers, Luc

APPLICANT: Filvaroff, Ellen

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P3330R1C30

CURRENT APPLICATION NUMBER: US/10/123,155

CURRENT FILING DATE: 2002-04-15

Prior Application removed - See Palm or File Wrapper

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 543

LENGTH: 3721

TYPE: DNA

ORGANISM: Homo Sapien

US-10-123-155-543

Query Match 48.4%; Score 2790.5; DB 14; Length 3721;
Best Local Similarity 39.7%; Pred. No. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623; Gaps 83;
QY 1 GGGGCC-----CATCT-----GTTTC-----AGCCAG-TGCCCAAG 31

Run on: May 7, 2004, 15:19:21 ; Search time 54 Seconds
(without alignments)
4759.756 Million cell

Perfect score: 5766
Sequence: 1 GGGGCCCATCTGTTTCAGC.....TTAGATGGTGAAAAAAA 926

Searched: 1140673 seqs, 277566755 residues

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Maximum Match 100%
Listing first 45 summaries

```

1: /csm2_6/ptodata/2/pubpaa/US06 PUBCOMB pep:
2: /csm2_6/ptodata/2/pubpaa/PCT NEW PUB pep:
3: /csm2_6/ptodata/2/pubpaa/US06 NEW PUB pep:
4: /csm2_6/ptodata/2/pubpaa/US06 PUBCOMB pep:
5: /csm2_6/ptodata/2/pubpaa/US07 NEW PUB pep:
6: /csm2_6/ptodata/2/pubpaa/PCTUS PUBCOMB pep:
7: /csm2_6/ptodata/2/pubpaa/US08 NEW PUB pep:
8: /csm2_6/ptodata/2/pubpaa/US09A PUBCOMB pep:
9: /csm2_6/ptodata/2/pubpaa/US09C PUBCOMB pep:
10: /csm2_6/ptodata/2/pubpaa/US09C PUBCOMB pep:
11: /csm2_6/ptodata/2/pubpaa/US09 NEW PUB pep:
12: /csm2_6/ptodata/2/pubpaa/US10A PUBCOMB pep:
13: /csm2_6/ptodata/2/pubpaa/US10B PUBCOMB pep:
14: /csm2_6/ptodata/2/pubpaa/US10C NEW PUB pep:
15: /csm2_6/ptodata/2/pubpaa/US10 NEW PUB pep:
16: /csm2_6/ptodata/2/pubpaa/US60 NEW PUB pep:
17: /csm2_6/ptodata/2/pubpaa/US60 PUBCOMB pep:
18: /csm2_6/ptodata/2/pubpaa/US60 PUBCOMB pep:

```

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2790.5	48.4	3721	12	US-10-142-426-543	Sequence 543, App	
2	2790.5	48.4	3721	14	US-10-123-155-543	Sequence 543, App	
3	2790.5	48.4	3721	14	US-10-146-731-543	Sequence 543, App	
4	2790.5	48.4	3721	14	US-10-140-472-543	Sequence 543, App	
5	2790.5	48.4	3721	14	US-10-141-761-543	Sequence 543, App	
6	2790.5	48.4	3721	14	US-10-142-885-543	Sequence 543, App	
7	2790.5	48.4	3721	15	US-10-158-790-543	Sequence 543, App	
8	2790.5	48.4	3721	15	US-10-137-971-543	Sequence 543, App	
9	2790.5	48.4	3721	15	US-10-140-923-543	Sequence 543, App	
10	2790.5	48.4	3721	15	US-10-141-756-543	Sequence 543, App	
11	2790.5	48.4	3721	15	US-10-141-759-543	Sequence 543, App	
12	2790.5	48.4	3721	15	US-10-140-805-543	Sequence 543, App	
13	2790.5	48.4	3721	15	US-10-140-864-543	Sequence 543, App	
14	2744.5	47.6	1942	12	US-10-142-426-515	Sequence 515, App	
15	2744.5	47.6	1942	14	US-10-123-155-515	Sequence 515, App	

Query Match 48.4%; Score 2790.5; DB 12; Length 3721;
Best Local Similarity 39.7%; Pred. NO. 1.3e-131;
Matches 609; Conservative 0; Mismatches 302; Indels 623;

```

Db      631 TGTTCCTCAATGACTCTCAACACTGTAACTGGAATTTGAAGCTATGCG--TATTCGAC 688
Qy      319 AGCACGTCATCGACTACATCAGG-GACCTTCAGTTGGAGCTGAACCTCGGAATCCGAATTT 377
Db      689 GGC-CGAAATCGAGTACAAATGGTGTACGTC-GAAGGAGCCGAATTTGTCAGACGCGTC 746
Qy      378 GGGACCCCGGGGGCGGAGGGGTGCGG-----GTC-----GSG 410
Db      747 AAGCCGAGCGCAACATCGAATGTGAGTTATTAATTCATAAATCTGCCAAAACGG 806
Qy      411 GCTCCGCTCAGCAC-----CC-----TCAACGG----- 433
Db      807 ACATTTGCCAGCACTTCATCGGGACCTACTCTCGICTACGGGTTAGTTTCATATTGAT 866
Qy      434 CGAGATCAGG-----CCCTGAGGCC----- 455
Db      867 CGCGA-CAGCGGCTTCTACTTCTTCAAAATATTTTCCCTGCGAGCTCGTCTAGTTT 925
Qy      456 -----GA-----GGCGGCAT-CCGTTTC-----TGC 475
Db      926 ATCATGATCTCATTCTGGATCAATCGTGACTCGGCGCTTCGGAAACCTAATCGGTAC 985
Qy      476 G--GACGAT-----CGCATCTTGTGTGCTGAAGCG-----CCTCCCCCAG 514
Db      986 GATGACGGTCTCACTGAGACTCATCTATGAC-CGGAAACCAATCGACGCTTCCACCAG 1044
Qy      515 GGACCGGCG-----GAC-----CC-----CAGCCATC--CAGG-----GGG 543
Db      1045 TTGCTATGTAAGAGCGTTGATGATTCCTCGGTTCTGCTATCTTGGTTATCTAG 1104
Qy      544 C-AAGAGGAATTACGTCTCTGTG-GGTCTCCCC-----AAGCGCCTCGCCGGA 592
Db      1105 CGTTGATCGAGTAGC--CCTGTGTTGCCTACTCAAAAAGAGAACGAGGATCGTCGGA 1161
Qy      593 TCTGAGGGAACAGAGCCGATCGGCGCC-CTGCGCC-----CTTAACGT 638
Db      1162 --GAAGAGAGAGAGAGAGCGGAGATAAACTGTCTCGCCGACACCTGATATCTTCACGA 1219
Qy      639 CATCCAGCCTGGGCTGAGGCTGAGGCACTGGCGAGGAGAGGCGC--TCTCTCTGCGAC 696
Db      1220 CGTCC-SCCT--TGCCGA-----ATGCACATGCAAGC--GGCTCCAACTCGATCATC 1268
Qy      697 ACC-----TACT-----AGTCAC-----CAGAGACTT--TAGGG 723
Db      1269 GCCGTATCAAGCAGTCGAATCGAATCTGTGTGTCAGTCACAGTCACATTTGACATCGTCAGC 1328
Qy      724 GGT---GGGATTCACCTCGT-----GTGTT---TCTATTTTGT----- 756
Db      1329 CGTCCGGGTTCTCTGTTTTCATCTGTTCACACACTCTCTCTGGCTGATTCCTACTG 1388
Qy      757 ---AAAAGCAGACATT-----TTAAAAATGGTCAG-----TTTGGTCTTC- 796
Db      1389 TACAAATCCAGCGTCTGCCGTATATTAGTGAACACGAGGAGACCGTTGCCGATCTCCA 1448
Qy      797 -----TCA--GATTTCTGAGAAATGCTTTGTATTGTATTAATTACATGATCAC----- 843
Db      1449 GACCTTCATTAATCTCAATCCAACTTCCTCATCATTTTCCATTTCGAATATCTCTTTTC 1508
Qy      844 -----CGACTGAGAAATGTTTACAAATAGTTCTGTGGGCTGTTTTTTTGTAT 894
Db      1509 TTGCACAGAGCCCTTTTTCGTTTTTTTATTTGATTTATTTTACGGATTTTAGATAA 1568
Qy      895 TAAACAAA-----TAATTAGATGGTGAATAAATAAATAA 926
Db      1569 TGACAGATGCTCATTCCTCAATAAATTTATTTTAAATGTCGAAAAAATAA 1622

```

Search completed: May 7, 2004, 15:20:45
Job time : 43 secs

Matches 541; Conservative 0; Mismatches 354; Indels 448; Gaps 77;
QY 7 CATCTGTTTACG---CAGT---CGCAGAAATCATGAAGTCGCGAGTGGCAGACG 59
Db 63 CAAGCTGGTGGCTTTTTCATCAACACCAAGAACCCAGTAAAGCAGCTGTCAGGACG 122
QY 60 ----GCCACCG 112
Db 123 CTGTGCTCACTGCAGAAAAGGGTGCACCTTAC-CATTTCTGCA--CAGGAAATGTAAAG 179
QY 113 CGGTG---CGGCGAGGTGGTGGCTGTCTGTGAGCAGAG--CGTGGCCATCTCG 164
Db 180 TGGTGTATTACCGGGGAC--TGTAACCTTTGAATC---CAGAGGCCATGATGAATCA 233
QY 165 CG---CTGCCCGGGGCG---CGGGGCG---CGCGTGGCTG---CCCTGCTGGA 205
Db 234 CTATCCAGCCAGGAGACATAGTCATGTGGTGAAGCCAACTGGAGAACCCGGCTGGC 293
QY 206 C--GAGCAGCA---GGTAAAC---GTGTGTCTTACGACATGAACGG--- 244
Db 294 TTGGAGGAGATTTAAAGGAAAGACAGGGTGGTTCCTTCCGCAACTATGACAGAAATCC 353
QY 245 CTGTATAC---TCACGGCTCA-AGGAGCTGGTGCC-----CA-----CCCTGCC 284
Db 354 CAGAAATAGGTTCCCGCTCCAGTGAACCACTGACTGATTCACATCTGCCCTGCC 413
QY 285 CAGAACCG--CAAG-GTGAG-----CAAGGTGAGAT---TCTCAG---CA 322
Db 414 CCAAACTGGCTTGGCTGAGACCCCGCCCTTTGGCAGTAACTCTTCAGAGCCCTCA 473
QY 323 CGTCATC-----GACTACATCAG-----GGACCTTCAGT-----TGGAGC 357
Db 474 CCAACCTTAATACTGGGCGGCTTCAGCTCCAGTGGCCCAACAGCAGGATGAAGAC 533
QY 358 TGAATCTGGAATCC---GAAG-TTGGACCCCGG-----GGGCGAGGGT- 400
Db 534 CAGAAACGATAACTGGGATGATGGGCGAGCCGCGCTCTCTCAACCGTTCAAGTGGCG 593
QY 401 GCGGTCGCGGC---TCGC-CTCAGACCTTCAACGCGCA---GATCAGCGCCCTGAC 451
Db 594 GCGAGTTAAGGAGAGGTCCGCTTTACTTCAGCCACGCGCCACTGCTC--CTCCCGCTC 651
QY 452 -----GGCC-----GAGGCGGCATGC-----GTTTCCT-- 473
Db 652 TCCTGTGCTAGGCGAGGTGAAAGGTGGAGGGGCTTCAAGCTCAAGCCCTATATCTTTG 711
QY 474 ----GC-----GGACATGCGATCTTGTGCTGTAAGC---CCTCCCG--CAGGGA 517
Db 712 GAGAGCCAAAAGACACACCTTAAATTTTAAACAAAATGATGTCAATCAGCGTCTGGA 771
QY 518 CCGGC-----GGA-----CCCCAG--- 531
Db 772 ACAGCAAGACATGTGGTGGTTGGAGAGTTCAAGTCAAGAGGTGGTTCCCAAGTC 831
QY 532 -----CCATCCAGGGGGGCAAGAGGAA--TTAGTGC-----TC 562
Db 832 TTACGTGAACCTCATTTTCAGGGGCCCATAGGAAGTCTACAAGATGGATTCTGTCTTC 891
QY 563 TGTGGG-----TCTCCC-----CCAA--CGCGCTCGCCCG 591
Db 892 AGAGAGTCTGTAGTCTAAGCGTAGCTCTCCAGCAGCAAGCCGAGCCAAAGCCG 951
QY 592 ----ATCTGAGGAGAACAGACCG-----ATCGCGGCGCACTGCGCCCTTAA 635
Db 952 TCGTTTCGGGAGAGAAATTTGCCAGGTATTATGCTTCATACACCGCCACCG--GCCCGAG 1010
QY 636 CTCG--ATCCAG--CTGG--GGCTGAGCTGA-----GGCACTGGCG--- 672
Db 1011 CAGCTACTCTCGCCCTGTGTAGCTGATTTGATCCGAAAGAACCCAGGTGATGG 1070
QY 673 -AGGAGAGGCGCTCTCT--CTC-TG-----CACACCT-----ACTAGTCAACAGACT 717
Db 1071 TGGGAGGAGAGCTGCAAGCAGCTGGGAAAGAGCCGACAGTAGGCTGTTCACAGTAAT 1130

QY 718 TTAG-----GGGTGGG--ATTCCACT-CGTGTGTTCTATTTTTGA 757
Db 1131 TATGTAAAGCTTCAAGCCCTGGGACGAGCAAAATCACTCCAACAGAGCCACCTAGTCA 1190
QY 758 AAGCA-----GACATTTTAAAAATGGTCACTTTGGTGTCTCT-----CAGATTT 804
Db 1191 ACAGCATTAGCGGACGTGTCCAGGTGATTTGGGAT--GTACGACTACACGCGCAGAATG 1248
QY 805 CTGAGGAAATTTGCTTT-----GTATTGTATATTACAATGA--TCACCG 845
Db 1249 ACATGAGCTGGCTTCAACAGGGCCAGATCATCAAGCTTCAACAGGAGGCCCTG 1308
QY 846 ACTG--AGATATTTTACAATAGTCTGTGGGCTGTTTTGTATTATTAACAAAT 903
Db 1309 ACTGGTGAAGAGAGAGT--CAATGGACAAGTGGGCTCTT-----CCCATCAAT 1358
QY 904 AATTAGATGCTGAAAAA 926
Db 1359 TATGT-GAAGCTGACACAGACA 1380

RESULT 14

US-09-627-650B-1
; Sequence 1, Application US/09627650B
; Patent No. 640872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.00093
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1652
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-1

Query Match 40.4%; Score 2327; DB 4; Length 1652;
Best Local Similarity 37.3%; Pred No. 8.9e-98;
Matches 579; Conservative 0; Mismatches 321; Indels 554; Gaps 84;

QY 1 GGGGC-CAATTC-----TGTTTCAAGCCAGTCCG-----CAAGAATCA 36
Db 95 GAAGCTCAATTCGAGAAATGGCTCGTCCATTCACTTATCGTACTCTCTCCGCACA 154
QY 37 TGAAGTCCGCA-GTGG-----CAG-----CACGCG-CACCGCG 69
Db 155 TCTGTGTCTCATGTGGTTGTGACACAGGATGAGGACTCATATCAACTCACTCTCT 214
QY 70 C-----CGCG-----GGCCCCAGCTCGCGCTG-----AAGGCG--G 99
Db 215 CTATCATGTTCTCGATAGACTCAGAAATCGCACTTATGATAAAAGATTACGGCCAG 274
QY 100 GCAAGACAGCGGCTGGCGGAGGTGG---TGGCTG-----TCTGTCTGAG 146
Db 275 GTATGGTGAAGAGC-----CAGTCGACGTTGGAAATTACGATACAGGTTTCTTCAATCTCTG 330
QY 147 CAG-----AGC-----GTGCGCATCTC-----GCG-CTGCGCGGGC--- 176
Db 331 CAGTTTCAGAAAGTTGATATGACCTTCACTTAGACTTCTATACATGCGTCAAAACGTTGCAAG 390
QY 177 -----GCC-----GGGG-----GGGCTGCTGCTGC 196
Db 391 ACCCTGACTAGGCTTCGGAAGTCTTGTATTGGGACTTTTCCAAAGAAATCGACTCTTA 450

QY 59 CGCC--ACGCCGCCCGCGGCC-----CCAG-----CTGCGCG-----CTGAAGCCGGCAA 103
Db 1194 CGCCGAACACAGCAGCGAAACGAGATGCCAATGTTCAACGCGAGCGCGAGCGCGCAA 1253
QY 104 -----GACA-----GGAG-----CGGTGCGGGGAGGTG-----GT 130
Db 1254 TAATAATTATACGAATGACATTTATGTCGAAAATTCGACGCGCTGCCAAAGCTATGT 1313
QY 131 GC--GCTGTG--TGT--CTGAGCAGAGC-----GTGGCCATCTC 163
Db 1314 ACAGGCTGACTGTACTTTGCGGACACAAATTCCTCTATGATCCATTCGATGAGATGCC 1373
QY 164 -----GGCG--TGCGG-----GGGCGCGG--GG 183
Db 1374 AGAAAAATTGATTGCGGAGCGAATTCMAATGATGCAACATCAACGCTTTGTCAAGAGCG 1433
QY 184 CGCGCTGC--CTGCGCTGCT--GGACGAGCAGCAGGTAA-----ACGTGTG 227
Db 1434 CGACATACGCTATGCGCGCTCCATTCGCGCGCGCGAAAAGGCTTCCAGACATGCTG 1493
QY 228 CTCTAGACATGAACGGCTG-----GTC-----CGGCTCC 415
Db 1494 C-CAACG--ATGGACGCGCTCAAAAAATCGAATAGCTTAGCCGATACGTTTCCCATTTGTC 1550
QY 248 TTAATCAGCGCTCAAGG-----AGCTGGT-----GC-----CCAC----- 277
Db 1551 TTTCTCTATCTTAAATATAGTCTACTGTTGTATATGAATATCTAAGCTTAACTGCTC 1610
QY 278 ---CCTGCCCGAG--AACCCAGAGTG--AGCA-----AGTGAGATTTCTCCAGCAC 323
Db 1611 GGCMAAGTCCAGGAGAACCAAGTGGCAGCAGATCCACTGATGCGTATTCGACGCG--C 1669
QY 324 GTGATCAGCTACATCAGG--GACCTCAGTTGAGAGCTGAACTCGGAATCGMAATTGGGAC 382
Db 1670 GAAATCAGTACAAATGGTGTAGCTC--GAAGGAGCGGATTTGTCAGACGGGTCAAGC 1728
QY 383 CCCCGGGCGCGGGCTGCCG-----GTC-----CGGCTCC 415
Db 1729 CGACGCGAACATCGAACTGTGAGTTATAAATCACTAAATCTGCCAAAACGAGACT 1788
QY 416 GCTCAGAC-----CC-----TCAACG-----CGAGA 438
Db 1789 TGCAGCAGCTTCATCGGGGACCTACTCTGCTACGGGTAGTTTATATTTGATCGCGA 1848
QY 439 TCAGCG-----CCTGACGGCC----- 455
Db 1849 -CAGGGCTTCTACTTCTTCAATATTTTCCCTGCGCAGCTCGTGTAGTTTATCAT 1907
QY 456 -----GA-----GGCGCAT--CGCTTC-----TGCG--GA 478
Db 1908 GGAATCTCATCTGGATCAATCGTACTCGGCGCTTCGGAACCTTAATCGGTACGATGA 1967
QY 479 CGAT-----CGCATCTTGTGCTGAAGCG-----CCTCCCGCAGGGACC 519
Db 1968 CGGTGCTCACTGAGCTCACTTATGAC--CGGAACCAATCGACGCTTCCACAGTTGCC 2026
QY 520 GGCG-----GAC-----CAGCCATC--CAGG-----CGGC--AAG 547
Db 2027 TATGTAAAAGCGTTGATGATTTCTCGGTTTCTGCTATCTCTGTTATATCTGCGGTG 2086
QY 548 AGGAATTACGTGCTGTG--GGTCTCCCC-----AACGCGCTCGCCGATCTGA 597
Db 2087 ATCAGTACG--CCTGTTGCTCTCTCAAAAAGAGAACGAGGATCGTCGA--GAA 2141
QY 598 CGGAGAACAGACCGATCGCGGCC--ACTGGGCC-----CTTAATCTGATCC 643
Db 2142 GAGAGAGAGACGAGCATAACTGCTCGCGCACACCTGATATTTCTCAGCAGTCC 2201
QY 644 AGCCTGGGCTGAGCTGAGCAGCTGGCGAGAGAGGCGG--TCCCTCTCTGACACCC-- 699
Db 2202 -GCCT--TGCCGA-----ATGCATGCAACGC--GGCTCCACCTGATCATCGCGT 2250
QY 700 -----TACT-----AGTCAC-----CAGAGACTT--TAGGGGT-- 726

Db 2251 CATCAAGCAGTCGAATCGATTCTGTGTGTCAGTCAAGTCACATTGACATCGTCAGCGTGC 2310
QY 727 -GGGATTCCTACTCGT-----GTGTT-----TCTATTTTTTG-----AA 758
Db 2311 CGCGTTTCTCTGTTTTCATCTTGTTCACACTCTCTCTGCTGCTGATCTACTGTACAA 2370
QY 759 AAGCAGACAT-----TTAAAAATGTCAGC-----TTTGGTCTTC----- 796
Db 2371 ATCCAGCGCTCGCGCTATATTAGTGAACAGAGGGTGACCGTTGCGATGCTCCAGACCT 2430
QY 797 TCA--GATTTCTGAGGAAATGCTTTGTTATTTATTTATTAACAATGATCAC----- 843
Db 2431 TCATTATCTCAATCCAACTTCTCATCAITTTCCATTTTGAATATCTCTTTTCTTGCA 2490
QY 844 ---CGACTGAGATATTGTTTACAATAGTCTCTGCGGCTGTTTTTTGTTATTAAAC 899
Db 2491 CAGAAGCGTTTTTTCGTTTTTTTATTTATTTATTTTACGATTTTATGATGATGCAC 2550
QY 900 AAA-----TAATTTAGATGTTGAAAAA----- 926
Db 2551 AGATGCTCATCTGCTCAATAAATTTATTTTAAATAAAAAA----- 2599

RESULT 13
US-08-630-915A-37
; Sequence 37, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, NO. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF INVENTIONS: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-630-915A-37

Query Match 40.5%; Score 2337.5; DB 4; Length 1400;
Best Local Similarity 40.3%; Pred. No. 2.7e-98;

; ORGANISM: Caenorhabditis elegans		US-09-627-650B-9	
Query Match		41.3%; Score 2383.5; DB 4; Length 2601;	
Best Local Similarity		38.0%; Pred. No. 3.6e-100;	
Matches 589; Conservative		0; Mismatches 313; Indels 647; Gaps 87;	
QY	2	GGGCC-----CATTCGTTC-----AGCC-----AGT- 24	
Db	1074	GGGCTGGATGTGTTCTTAAATTTTTCGTAATGGTATTCGCTCGTTCGAGTA 1133	
QY	25	CGCCA-AGAAT-----CATGAA-AGTCG-----CGAGTGGCAGCA-----C 58	
Db	1134	CGCCATAGTATCTTACATGAATAAAGACTGGTCTCCGACGGGAAAAACGAGAAAGC 1193	
QY	59	CGCC--ACCGCCGCGCGGCC-----CCAG-----CTGGCG--CTGAAGCGCGCAA 103	
Db	1194	CGCGAACACACGACGAGAACAGATGCCAATGTTCAACGCGAGCGCCGAGCGCCAA 1253	
QY	104	-----GACA-----GCAG--CGTCCGGCGGAGGTG-----GT 130	
Db	1254	TAAATATTCATACGAATGACACTTATGTGCAAAATTCGACGCTGCCAAAAGCTATGT 1313	
QY	131	GC--GCTGTC-TGT-CTGACGAGC-----GTGGCCATCTC 163	
Db	1314	ACAGGCTGACTTGTACTTTTCGCGACACAAATCTCTATGAATCCATGATGGAGATCC 1373	
QY	164	-----GCGC-TGCCG-----GCGCCGCGG--GG 183	
Db	1374	AGAAATGTGATTCGCGGACGATTCCAATGATGCAACATCCACGCTTGTGCACAGCG 1433	
QY	184	CGGCCCTGC--CTGCCCTGCT--GGACGAGCAGCGTAA-----ACGTGCTG 227	
Db	1434	CGCACATACGCTATGCGCGGCTCAATTCGCGCGCGGCGAAGGCTTCCAGACATGCTG 1493	
QY	228	CTCTACGACATGACGCGCTG-----GTC-----GTC-----CGGCTCC 415	
Db	1494	C-CRACG--ATGGAGCGCTCGAAATCGATAAGCTTAGCCGATACGTTTCCATGTC 1550	
QY	248	TTACTACGCTCAAGS-----AGCTGCT-----GC-----CCAC--- 277	
Db	1551	TTTCTCTATCTTCAATATAGTCTACTGTTGTATATGAATAATCTAAGCTTAACTCGTC 1610	
QY	278	---CCTGCCCGAG--AACCCGACAGGTG--AGCA-----AGTGGAGATTTCTCCAGCAC 323	
Db	1611	GGACAGATCCAGGAGAACGACAGTGGCAGCAGATCCACTGATGCGTATTCGACGCG-C 1669	
QY	324	GTATCGACTACATCAGG-GACCTTCAGTTGGAGTGAATCGGAATCCGAATGGAGTGGAC 382	
Db	1670	GAATTCGAGTACAAATGGTGTACGTC-GAAGGAGCGCAATTTGTCGACAGCGGTCAAGC 1728	
QY	383	CCCCGGGGCGGAGGCTGCCG-----GTC-----GTC-----CGGCTCC 415	
Db	1729	CGACCGAATCGAATGTGAGTTATAATTCATAAATCTGCCAAAACGGACACT 1788	
QY	416	GCTCAGCAC-----CC-----TCAACGG-----CGAGA 438	
Db	1789	TGCCAGACATTCATCGGGGACCTACTCTCGTACGGGTAGTTTATATTCATTCGCGA 1848	
QY	439	TCAGCG-----CCTGACGCGC-----GTC-----GTC-----CGGCTCC 415	
Db	1849	-CAGCGGTTCTACTTTCTTCAATATTTTTCCTGCCAGCTTCGTCGTAAGTTTATCAT 1907	
QY	456	-----GA-----GCGGGCAT-GCGTTCC-----TCGCG--GA 478	
Db	1908	GGATCTCATCTGGATCAATCGTACCTCGGCGCTTCGCGAACCTTAATCGGTACGATGA 1967	
QY	479	CGAT-----CGATCTTGTGCTGAAGC-----CCTCCCCAGGAGC 519	
Db	1968	CGGTGCTCACTGAGACTCATTTATGAC-CGGACCAATACGAGTCTTCCACAGTTGCC 2026	
QY	520	GGCG-----GAC-----CC-----CAGCCATC--CAGG-----GGGC-AAG 547	

RESULT 12		
US-09-436-063C-9		
; Sequence 9, Application US/09436063C		
; Patent No. 6407210		
; GENERAL INFORMATION:		
; APPLICANT: Bamber, Bruce		
; APPLICANT: Jorgensen, Erik		
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and		
; TITLE OF INVENTION: Methods Related Thereto		
; FILE REFERENCE: P-1095Corrected		
; CURRENT APPLICATION NUMBER: US/09/436,063C		
; CURRENT FILING DATE: 1999-11-08		
; PRIOR APPLICATION NUMBER: 60/107727		
; PRIOR FILING DATE: 1998-11-09		
; NUMBER OF SEQ ID NOS: 18		
; SOFTWARE: PatentIn Ver. 2.1		
; SEQ ID NO 9		
; LENGTH: 2601		
; TYPE: PRT		
; ORGANISM: Caenorhabditis elegans		
US-09-436-063C-9		
Query Match		
Best Local Similarity		
Matches 589; Conservative		
0; Mismatches 313; Indels 647; Gaps 87;		
QY	2	GGGCC-----CATTCGTTC-----AGCC-----AGT- 24
Db	1074	GGGCTGGATGTGTTCTTAAATTTTTCGTAATGGTATTCGCTCGTTCGAGTA 1133
QY	25	CGCCA-AGAAT-----CATGAA-AGTCG-----CCAGTGGCAGCA-----C 58
Db	1134	CGCCATAGTATCTTACATGAATAAAGACTGGTCTCCGACGGGAAAAACGAGAAAGC 1193

```

; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 13
; LENGTH: 2088
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-551-853D-13

Query March 41.4%; Score 2385; DB 4; Length 2088;
Best Local Similarity 37.6%; Pred. No. 2.7e-100;
Matches 600; Conservative 0; Mismatches 312; Indels 682; Gaps 93;

QY 1 GGGCCCATCTCT-GTTTQAG-CCAGT-----CGCCAGA-ATCATGAAG-TCG 45
DB 373 GATGCCC-TTCTGTCCTGCAAGTGGCAATTTTACACAGGAGGATGGATGTTG 431
QY 46 CCAGTGGCAGACCCG-----CACCGCGCGCGCGCGCGCTGCGCTGAAGGCGG- 99
DB 432 CGAAATCTCATCTTCACTGGCACAACCGTCCCAAGAGAGATGCACTGAGAAGAGTACCA 491
QY 100 ---GCAAGACAGG-----AGCGTGGGG-----CGAGTG----- 128
DB 492 CTTGCATGACTACCGCATGTTGCTGCCCTGCGGAATTCACAAGTTCGAGGGGTAGATT 551
QY 129 ---GTG-----CGCTGTCTG-----TCTG-----AGCAGAGCGTGG 156
DB 552 TGTGTGTTGCCACTGGCTGAAGAAAGTGACAAATGTGATTTCTGCTGATGGGAG-GAGG 610
QY 157 CCATCTCGCGCTG-CGCGGCGCGCGCGCGCGC-CTGCGTCCCTGCTGGAGC----- 207
DB 611 ATGACTCG-GATGTCTGTGTGGCGGAGCAGACAGACTATGCAATGGGATGGAGTGAAGAC 669
QY 208 -----AGCAGC--AGGTAAAGT-GCTG-----GCTGTCTACTCAC----- 255
DB 670 AAAGTGTAGTAGTAGCAGAGGAGGAGGAGTGGTGGAGTGGAGAGAGGAGGAGGAGGAG 729
QY 232 -ACGACATGAAC-----GCTGTCTACTCAC-----GCTGTCTACTCAC----- 789
DB 730 GATGACAGGACGATGAGTGTGATGAGTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAG 799
QY 256 ---GCTCAAGAG-----CTGG---TGCCCACTCTGCG-CCAGAACCGGA-AG---GTG 299
DB 790 GAAGCCACAGAGAGAACCAACAGCATTTGCCACCAACCAACCAACCAACCAACCAACCA 849
QY 300 AGCAAGGTG---GATTTCT-----CCAGCAGC-----GTCATCGAC---TAC 335
DB 850 GAAGAGTGTGTCAGTTCCTTACACAGCAGCAGCAGTACCCCTGATGCGGTTGACAGTAT 909
QY 336 ATC-AGGACCTTCACTTGGAGCTGAAC-----TC--GGAATCCGAAG-----TT 377
DB 910 CTCGAGACACCTGGGATGAGATGAACATGCCATTTTCCAGAAAGGAGGAGGAGGAGG 969
QY 378 GGGACC-----CCCGGG-----CCCGGG-----CCCGGG----- 389
DB 970 GAGGCCAAGCAACGAGAGAGATGTCCAGGTCTATGAGAGATGGAGAGGAGGAGAGCT 1029
QY 390 -----GGCCGAGGGCTG-----CCCG-----CCCG-----CCCG----- 411
DB 1030 CAAGCAAGAACTTGCCTAAAGCTGATAGAGAGGAGGAGTATCCAGCATTTCCAGGAGAA 1089
QY 412 -----CT-----CCCG-----CCCG-----CCCG-----CCCG----- 429
DB 1090 GTGGAATCTTGGAAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1149
QY 430 -----ACGGGAGATGAG-----CGCC-CTGAC-GGCGGAGG----- 459
DB 1150 GCAGAGTGAAGCCATGCTCAATGACCGCGCGCGCTGGGAGAACTACATCAACC 1209
QY 460 ---CGCGATCG-GTTCCTCGGACGATGCGATCTTGTGT-----CGCTGAA----- 501

```

RESULT 11

```

US-09-627-650B-9
; Sequence 9, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT

```

```

DB 1210 GCTGTGAGGCTGTCTCTCC--TCGGCTCGTCACTGTTTCAATATGCTAAAGAGATG 1267
QY 502 ---GGC-----CTCCCCAGGAGC-----GG 521
DB 1268 TCCGCGCAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1327
QY 522 CGGACCCC-----AGCC-----ATCCAGG-----GGCA-----AGAG-GAATT 554
DB 1328 TGGATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1387
QY 555 AGCTGC-----TCTGTGGGTCTCCC-----CCAACGCGCT-C-----GCCG----- 590
DB 1388 ATGAGCGATCAATCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1447
QY 591 -----GATCTG-----AGGAGAGAGAA-----CA 609
DB 1448 TTCAGGATGAAGTGTGATGAGCTGCTTCAGAAAGAGAGAGAGAGAGAGAGAGAG 1507
QY 610 CC-----GATCGG-----CGGCCAC--TGCGC----- 629
DB 1508 CCAACATGATTAGTGAACCAAGGATCAGTTACGGAACAGATGCTCTCATGCCATCTTTGA 1567
QY 630 CTTAACTGCTATCCAGCTGGGCT-----GAGGCTGAGG-----CA-CTGG----- 670
DB 1568 CCGAAACGAAACCAACCGCTGAGCTCCTCCCGTGAATGGAGAGTTGAGCTGGACGATC 1627
QY 671 -----CGAGGAG-----AGGCGCT--CCTCTCTGCACACCTACT-AGTCACC-----AG 712
DB 1628 TCCAGCCGTGGCATCTTTTGGGGCTGACTCTGTGCCAGCAACAGAGAAACGAAAGTTG 1687
QY 713 AGCTTTAGGGG-----GTGGGAT-TCCACTCG-----TGCT 744
DB 1688 AGCTGTGTATGTCGCGCTGCTGCCAGCAGAGATGACCTCGACAGGTTCTGGT 1747
QY 745 TCTATTTTTTGAAGAGAGACATTTTAAAAATGTCAGTTTGTGTC-----TTCTCAG 800
DB 1748 TGACAAATATCAAGACGAGAGAGATCTCTGAAGTG---AAGATGATGAGAGATTCGAG 1804
QY 801 ATTTCTGAGGAGAAAT--TGCTTTGTTGTTATATTACAAATGATCAACGACTGAGAAATTTG 858
DB 1805 ATGACTCAGGATATGAGTTCAATCAAAAATTT--GGTGTCTTTTG-CAGAGAGATG 1861
QY 859 TTT--TACATAGTTC-----TGTTGGGCTGTTTTTT-----TGT 891
DB 1862 GTTCAAAACAAAGGTGCAATCATTTGGACTCATGTTGGGCGGTGTTGTCTATAGCAGATGA 1921
QY 892 TATTAACAAATAATTTAGATGTTGAAAAA 925
DB 1922 TCTTCATCACCTTGGT--GATGCTGAGAGAGAA 1953

```


;; CURRENT APPLICATION NUMBER: US/09/548,367D
;; CURRENT FILING DATE: 2000-04-12
;; PRIOR APPLICATION NUMBER: US 60/155,493
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: US 09/404,133
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: PCT/US99/20881
;; PRIOR FILING DATE: 1999-09-23
;; PRIOR APPLICATION NUMBER: US 60/101,594
;; PRIOR FILING DATE: 1998-09-24
;; NUMBER OF SEQ ID NOS: 73
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 13
;; LENGTH: 2088
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; US-09-548-367D-13

Query Match 41.4%; Score 2385; DB 4; Length 2088;
Best Local Similarity 37.6%; Pred. No. 2.7e-100;
Matches 600; Conservative 0; Mismatches 312; Indels 682; Gaps 93;

QY 1 GGGGCCCATCTCTGTTTCAG-CCAGT-----CGCCAAGA-ATCATCAAG-TCG 45
Db 373 GATGCCCT-TTCTCGTCTGACAAAGTGCAAAATCTTACACAGGAGAGATGATGTTG 431

QY 46 CCAAGTGGCAGCACCGC-----CACCGCGCGCGGCCCCAGCTGCGCTGAAGCGCG- 99
Db 432 CGAACTCACTTTCACTGGCAGCACCGCTGCCAAAGAGACATGCAAGTGAAGAGTACCAA 491

QY 100 ---GCAAGACAGG-----AGCGGTGCGG-----CGAGGTG----- 128
Db 492 CTTGTCATGACTACGGCATGTTGTGCTGCCCTGCGGAATTGACAAGTCCGAGGGGTAGATT 551

QY 129 ---GTG---CGCTGTCTG-----TCGTG-----ACGAGAGCGTGG 156
Db 552 TGTGTGTTGCCACTGGCTGAAGAACTGACATGTGGATCTGCTGATGCGGAG-GAGG 610

QY 157 CCATCTCGCGGTG-CGCGGCGCGCGGCGCGC-CTGCGTCCCTGCTGAGC----- 207
Db 611 ATGACTCG-GATGTCTGTGTTGGCGGCGGAGCAGACAGACTATGACATGGGAGTGAAGAC 669

QY 208 -----AGCAGC--AGGTAAAGCT-GCTG-----CTCT 231
Db 670 AAGTAGTAGAAGTAGCAGAGGAGGAAGAGTGGCTGAGTGGGAAGAGAGCGGAT 729

QY 232 -ACGACATGAAC-----GGCTGTACTCAC----- 255
Db 730 GATCAGCAGGACGATGAGGATGAGTGTGATGAGTGAAGAGAGGCTGAGGAACCTTACGAA 789

QY 256 ---CCCTCAAGGAG-----CTGG---TGCCACCCCTGCC-CCAGAACCGCA-AG---GTG 299
Db 790 GAAGCCACAGAGAGAACACACAGCAATTGCCAACCAACCAACCAACCAACAGAGTCTGTG 849

QY 300 AGCAAGGTG-----GAGATTCT-----CCAGCAC-----GTCAATCGAC---TAC 335
Db 850 GAAGAGGTGTTTCAGTTCTTACAAACAGCAGCAGTACCCCTGATGCCGTGACAAAGTAT 909

QY 336 ATC-AGGACCTTCAGTTGGAGTGAAC-----TC---GGATCCGAG-----TT 377
Db 910 CTCGAGACACCTGGGGATGAGATGAATGCAATGCCCATTTCCAGAAAGCCAAAGAGAGCTT 969

QY 378 GGGACC-----CCCGGG----- 389
Db 970 GAGGCCAAGCAGCAGAGAGAAATGTCCTCCAGGTCAATGAGAAATGGGAAGGAGGAGAGCT 1029

QY 390 -----GGCCGAGGGCTG-----CCGG-----TCGGG----- 411
Db 1030 CAAGCAAGAACTTGTCTTAAGCTGTATAAGAGGAGTATTCAGAGATTTCCAGGGAGAA 1089

QY 412 -----CT-----CCGC-----TCAGC-----ACCTTCA-- 429
Db 1090 GTGGAATCTTTGGACAGGAAGACGACCAACGAGAGACAGCAGCTGTTGGAGACACATG 1149

RESULT 10

US-09-551-853D-13

; Sequence 13, Application US/09551853D

; Patent No. 6500667

; GENERAL INFORMATION:

; APPLICANT: GURNEY ET AL.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 29915/6280L

; CURRENT APPLICATION NUMBER: US/09/551,853D

; CURRENT FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 09/404,133

; PRIOR FILING DATE: 1999-09-23

QY 430 -----ACGGCAGATCAG-----CGCC-CTGAC-GGCCGAGG----- 459
Db 1150 GCAGAGTGAAGCCATGCTCAATGACCGCGCGCTGGCCCTGGAGAACTATACACC 1209

QY 460 ---CGGCATGC-GTTCTCGGAGCGATGCGATCTTGTGT-----CGCTGAA----- 501
Db 1210 GCTCTGAGGCTGTTCTCTCC-TCGGCCTGCTACGCTGTTCAATATGCTAAAGAGTATG 1267

QY 502 ---GCGC-----CTCCCCCAGGAGC-----GG 521
Db 1268 TCCGCGAAGAACAGAGGACAGACAGCACCTTAAGCATTTTCAGCATGTGGCATGG 1327

QY 522 CGGACCCC-----AGCC-----ATCCAGG-----GGCA-----AGAG-GAATT 554
Db 1328 TGGATCCCAAGAAAGCGCTCAGATCCGGTCCCGAGTTATGACACACCTCCGTGTGATTT 1387

QY 555 ACGTGC-----TCTGTGGGTCTCC-----CCACGCGCCT-C-----GCCG----- 590
Db 1388 ATGAGCGCATGAATCACTCTCTCTCTCCCTGCTCTACACGCTGCTGACGTGCCGAGAGA 1447

QY 591 -----GATCTG-----AGGAGAACAA-----GA 609
Db 1448 TTCAGGATGAAGTTGATGAGCTGCTTCAGAAAGAGCAAAACTATTTCAGATGAGCTCTGG 1507

QY 610 CC-----GATCG-----CGGCCAC-----TGCGC----- 629
Db 1508 CCAACATGATGATGAACAAAGGATCAGTTACGAAACGATGCTCTCATGCCATCTTTGA 1567

QY 630 CTTAACTGCATCCAGCTGGGCT-----GAGGCTGAG-----CA---CTGG----- 670
Db 1568 CCGAAACGAAACCAACCGTGGAGCTCTCCCGTGAATGAGAGTTTCAGCTGGACGATC 1627

QY 671 -----CGAGAG-----AGGGCGCT-CTCTCTGACACCTACT-AGTCACC-----AG 712
Db 1628 TCCAGCGCTGGCATCTTTTGGGCTGACTCTGCGCAGCAACACAGAAACAAAGTTG 1687

QY 713 AGACTTTAGGG-----GTGGGAT-TCCACTCG-----TGTGT 744
Db 1688 AGCTGTTGATGCCCGCCTGCTGCCGACCGAGGACTGACCACTCGACAGGTTCTGGGT 1747

QY 745 TTCTATTTTTGAAAAGCAGACATTTTAAAAATGTCACGTTTGGTGC-----TTCTCAG 800
Db 1748 TGACAAATATCAAGACGGAGGAGATCTCTGAAGTG---AAGATGATGCAAAATTCGAC 1804

QY 801 ATTTCTGAGGAAAT-TGCTTTGATTGTATTATCAATGATCACCAGCTGAGAAATATG 858
Db 1805 ATGACTCAGGATATGAAAGTTCAATCAAAATTT---GGTGTCTTTG-CAGAAATGTTG 1861

QY 859 TTT--TACAATGATTC-----TGCGGGCTGTTTTT-----TGT 891
Db 1862 GTTCAAAACAAAGGTGCAATCAATGGAATCATGCTGCTGGCGGTGTTGTCTATGACGACAGTGA 1921

QY 892 TATTAACAAATAATTTAGATGTTGAAAAA 925
Db 1922 TCTTCATCACCTTGT--GATGCTGAAGAGAAA 1953

Db 2303 CTACTGTACAAATCCAAAGCGTCTGCCGTATATTAGTGAACAACGAGGGTGAACGTTGCGAT 2362
Qy 792 GCTTC-----TCA--GATTCTGAGGAAATGCTTTGTTATGTTATATATTACAAATGATCAC 843
Db 2363 GCTCCAGACCTTCATTAAATCTCAATCCAACTTCCCATCATTTCCATTTGGAATATCTC 2422
Qy 844 -----CGACTGAGAAATATGTTTACATAGATCTCTGCGGCTGTTTTTT 888
Db 2423 TTTTCTCTGCACAGAAGCCTTTTTCGTTTTTTTTTTTATGATTTATTTTACGGATTTTT 2482
Qy 889 TGTATTAAACAAA-----TAATTTAGATGTTGAAATAAAAAA 926
Db 2483 AGATAATGCACAGATGCTCATTTGCTCAATAAATTTATTTTAAATAAAAAA 2542

RESULT 7
US-09-436-063C-3
; Sequence 3, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-10950corrected
; CURRENT APPLICATION NUMBER: US/09/436, 063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-3

Query Match 41.6%; Score 2397; DB 4; Length 2544;
Best Local Similarity 38.7%; Pred. No. 8.9e-101;
Matches 580; Conservative 0; Mismatches 314; Indels 606; Gaps 83;
Qy 2 GGGCC-----CATTCTGTTTC-----AGCC-----AGT- 24
Db 1074 GGGCTGCGATGTTTCTTAATTTTGTTCGTATGTTGCTGCTCGATGTA 1133
Qy 25 CGCCA-AGAAT-----CATGAA-AGTCG-----CCAGTGGCAGCA-----C 58
Db 1134 CGCCATAGTATCTTACATGAATAAAGACTGGTCTCCGACGGGAAAGAAAGC 1193
Qy 59 CGCC--ACCGCCCGCGGGCC-----CCAG-----CTGCGG-----CTGAAGGCG----- 99
Db 1194 CGCCGAACAACAGCAGCGAAACAGATGCAATGTTTCAACGCGAGCCGAGCGCCAA 1253
Qy 100 -----GCAAGACA-----GCGAG 112
Db 1254 TAATAATGCTGACTGTATCTTTGCGGACACAAATTCCTCTATGATCCATTTGATGGAGAT 1313
Qy 113 C-----GGTG-----CGGCGA-----GGTGCGCTGTC-----TGCTGAG- 146
Db 1314 CCAGAAATTTGATTTGCGGCGGATTCCAATGATGCAATCCACGCTTTGTACAGA 1373
Qy 147 CAGAGC-----GTGCGCTC-----TCGGCTGCG-----GGCG--GCGGGGCGG 187
Db 1374 CGCGGACATACGCTATGCGCGCTCCATTCGCGCGCGGAAAGGCTTCCAGACATG 1433
Qy 188 CTGCTGCTC-----CTGC-----TGACGAGC--AGCAGGTA-----AACGT 223
Db 1434 -CTGCCAAGATGGAGCGCTCGAAAAATCGATAAGCTTAGCCGATACGTTTCCCATG 1492
Qy 224 GCTGCTCTA-----CGACATGACGGCTG-----TTACTCAGC 257
Db 1493 CTTTCTCTATCTTCAATATAGTCTACTGTTGTATATGAAATATCTAAGCTTAAACTCGT 1552

Qy 258 C--TCAAAGGAGCTGGTCCCAACCT---GCCCCAGAACCCGAAAGGTGAGCAAGGTGAGA 312
Db 1553 CGGACAAAGATCCAGGAAACGACAAAGTGGCAGCAGATCCAC-----TGATGCGTA 1602
Qy 313 TTCTCCAGACGTCATCGACTCATCAGG--GACTTCAGTTGAGGCTGAGCTCGAATCC 371
Db 1503 TTCAAGCGC--CGAAATCGAGTCAAAATGGTGTAGCTC--GAAGGAGCCGAATTTGTCACA 1660
Qy 372 GAAGTTGGAGCCCGCGGGCGCGAGGGCTGCGG-----GTC-- 407
Db 1661 GCGGTCAAGCGCGAGCGGAACATCGAACTGTGAGTTATAAATTCATAAAATCTGCCAA 1720
Qy 408 ---CGGCTCGGTCAAGC-----CC-----TCAACGG----- 433
Db 1721 AACGGACACTGTCACGACTTTCATCGGGACCTTCTCTCGTCTACGGGTAGTTTCATA 1780
Qy 434 -----CGAGATCAGCG-----CCCTGACGCGC----- 455
Db 1781 TTTGATCGGA--CAGCGGCTTCTACTTTCTTCAATATTTTCCCTGCGACGCTCGTCGT 1839
Qy 456 -----GA-----GGCGGCAT--GGTTCC----- 472
Db 1840 AGTTTATATGATCTCAATTCGTGATCAATCGTGACTCGCGCCTTCGCGAAACCTTAAT 1899
Qy 473 ---TGCG--GACGAT-----CGCATTTGTCTGCTGAAGCG-----CCTC 508
Db 1900 CGGTACGATGACGGTGTCTACTGAGACTCATTTATGAC--CGGAAACCAATCGACGCTTC 1958
Qy 509 CCCAGGAGCGGGG-----GAC-----CC-----CAGCCATC--CAGG-- 540
Db 1959 CACCAATTGCTATGTAAGAGCCGTGATGATTTCTCTCGGTTCTCTGATCTCTCTGTTA 2018
Qy 541 ---GGG--AAGAGGAAATTAGTGTCTGTG--GGTCTCCCC-----AACGCGCTC 586
Db 2019 TACTGGGTTGATCGAGTAGC---CCTGTGTTCCCTACTCAAAAAGAGAGAGATC 2075
Qy 587 GCGGATCTGAGGAGAAACGAGCCGAGCGGCGC--ACTGCGC-----CT 632
Db 2076 GTCGGA--GAAG 2133
Qy 633 TAATGTCATCAGCTGGGCTGAGGCTGAGGCTGAGGCTGAGGAGAGAGAGAGAGAGAGAG 690
Db 2134 TCACGAGTCC--GGCT--TGCCGA-----ATGCATGCAACGC-----GGCTCAACCTCG 2182
Qy 691 CTGCACACC-----TACT-----AGTCAC-----CAGAGACTT-- 718
Db 2183 ATCATCGCTCATCAAGCAGTCCGATTCGATTCGATTCAGTCAAGTCAATGACATC 2242
Qy 719 -TAGGGGT---GGGATCCACTCGT-----GTGTT-----TCTATTTTTC----- 756
Db 2243 GTCAGCGCTGCGGCTTCTCTGTTTTCATCTTGTCAACACTCTCTTCTCGGCTGATT 2302
Qy 757 -----AAAAAGCAGACATT-----TTAAAAAATGGTCAAG-----TTTGGT 791
Db 2303 CTACTGTACAAATCCAAAGCGTCTGCCGTATATTAGTGAACACGAGGGTGACCGTTGCGAT 2362
Qy 792 GCTTC-----TCA--GATTCTGAGGAAATTCGTTTGTATGTTATATTACAAATGATCAC 843
Db 2363 GCTCCAGACCTTCATTAAATCTCAATCCAACTTCCCATCATTTTCCATTTGGAATATCTC 2422
Qy 844 -----CGACTGAGAAATATGTTTACAAATAGTTCTGTGGGCTGTTTTTT 888
Db 2423 TTTTCTTGCACAGAAGCCTTTTTCGTTTTTTTTTTTATGATTTATTTTACGGATTTTT 2482
Qy 889 TGTATTAAACAAA-----TAATTTAGATGTTGAAATAAAAAA 926
Db 2483 AGATAATGCACAGATGCTCATTTGCTCAATAAATTTATTTTAAATAAAAAA 2542

RESULT 8
US-09-548-372D-13
; Sequence 13, Application US/09548372D

QY 623 -ACTGGCC-----CTTAAGTGCATCCAGCTGGGCTGAGGTGAGCACT 668
DB 2074 TGCTCGCCGACACTGATTAATTTTCAGAGTCC-GCCT--TGCCGA-----ATGCACA 2125
QY 669 GCGAGGAGAGGGGCGC--TCCTCTCTGCACACC-----TACT-- 703
DB 2126 TGCACGCG--GCCTCCAACTCGATCATCGCGTCATCAACAGCAGTCGAATCGATTCTGT 2182
QY 704 ---AGTCAC-----CAGAGACTT--TAGGGGT--GGGATTCCACTCGT-----GTG 743
DB 2183 GTCAGTCAAGTCAATGACATCGTCAGCGGTGCGGGTTCCTCTGTGTTTCACTTGT 2242
QY 744 TT-----TCATTTTGTG-----AAAAGCAGCAATT--TTAAAAAT 778
DB 2243 TTCAACACTCTCTTCTGGTGTGATTCTACTGTACAAATCCAAAGCGTCTGCCGTATATTAGT 2302
QY 779 GGTCAAG-----TTTGGTCTTC--TCA--GATTTCTGAGGAAATGCTT 819
DB 2303 GAACACGAGGTGACCGTTCGATGCTCCAGACCTTCATTAATCTCAATCCACTTCCTC 2362
QY 820 TGTATTGTATATACAAATGATCAC-----CGACTGAGAAATATTGTTTAC 864
DB 2363 ATCATTTTCCATTTCCGATATCTCTTTTCTTGACAGAGCCCTTTTTCGTTTTTTTT 2422
QY 865 AATAGTTCGTGGGCTGTTTTTTTGTATTAAACAAA----- 902
DB 2423 ATGATTTATTTTACGGATTTTATAGATAATGCACAGATGCTCATTTGCTCAATAAATT 2482
QY 903 TAATTAGATGTTGAAAAA 926
DB 2483 TATTTTAAATTAATAAAAAA 2506

RESULT 6

US-09-627-650B-3
; Sequence 3, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE REFERENCE: 21101.0009U3
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR FILING DATE: 1999-11-08
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-3

Query Match 41.6%; Score 2397; DB 4; Length 2544;
Best Local Similarity 38.7%; Pred. No. 8.9e-101;
Matches 580; Conservative 0; Mismatches 314; Indels 606; Gaps 83;
QY 2 GGCC-----CATTCCTTTC-----AGCC-----AGT- 24
DB 1074 GGCTCTGGATGTTTCTTAATTTTGTGTAATGTATTCGCTCGTTGCTGGAGTA 1133
QY 25 GCGCA-AGAA--CATGAA-AGTCG-----CCAGTGGCAGCA-----C 58
DB 1134 CGCCATAGTATCTACATGAATAAACGACTGTCCTCCGCGGAGAAACGAGAAAGC 1193
QY 59 CGCC--ACCGCGCGCGGGCC-----CCAG-----CTGCGCG--CTGAGCGCG-- 99
DB 1194 CGCCGACAAACAGCAGCAAGATGCCAATGTTTCAACCGGAGCGCGGCGCCAA 1253

QY 100 -----CGAGACA----- 112
DB 1254 TAATAATGCTGACTTGTACTTTGCCGACACAAATTCCTCTATGAATCCATTTGATGAGAT 1313
QY 113 C-----GGTG-----CGGGCGA-----GGTGTGCGGTGTC-----TGTCTGAG- 146
DB 1314 CCCAGAAAAATGTTGATTGCCGACGATTCCTCAATGATGCAACATCCACGCTTCTGTACACA 1373
QY 147 CAGAGC-----GTGGCCATC-----TCGCGCTGCCG-----GGCG--GCGGGGCGCG 187
DB 1374 CGGGCAGATACGCTATGCGCGCTCCATTCGCGCGCGCGAAMAAAGGCTTCCAGACATG 1433
QY 188 CTTGCTGTC-----CTGCG-----TGACGAGC--AGCAGGTA-----AACGT 223
DB 1434 -CTGCCAACGATGAGACGCTCGCAAAATCGATAAGCTTAGCCGATACGGTTTCCCATTTGT 1492
QY 224 GCTGCTCTA--CGACATGAACGGGTG-----TTACTCACGC 257
DB 1493 CTCTCTCTAATCAATATAGTCTACTGTTGTATGAAATATCTAAGCTTTAACTCTGT 1552
QY 258 C--TCAGAGAGCTGGTCCCACT--GCCCGAGAACCGCAAGGTGAGCAAGGTGGAGA 312
DB 1553 CGGCAAGATCCAGGAGAGACGCAAGTGGCAGCAGATCCAC-----TGATGCGTA 1602
QY 313 TTCTCCAGACGCTCATCGACTACATCAGG--GACCTTCAGTTGGAGCTGAATCGGAATCC 371
DB 1603 TTCCAGCGC--CGAATCGACTACAAATGTTGTACGTC--GAGGAGCGGAATGTTTCGACA 1660
QY 372 GAAGTTGGGACCCCGGGGGCGGAGGCTCCCG-----GTC-- 407
DB 1661 CGGTCAAGGCGGACGCGAARCTCGAACTGTGAGTTATAAATTCCTAAATCTGCCAA 1720
QY 408 ---CGGGCTCCGCTCAGCAC-----CC-----TCAACGG----- 433
DB 1721 AAACGAGACTTTGCCAGCACTTCATCGGGGACCTACTCTCGICTACGGGTAGTTTCATA 1780
QY 434 -----CGAGATCAGCG-----CCCTGACGGCC----- 455
DB 1781 TTTGATCGCA--CAGCGGCTTCTACTTTCTCAAAATATTTTCCCTGCGAGCTCGTCTGT 1839
QY 456 -----GA-----GGCGGCAT--CGCTTCC----- 472
DB 1840 AGTTTATATGATCTCATCTGATCAATCGTACTCGGCGCTTCGCGAACCTTAAT 1899
QY 473 ---TCGCG--GACGAT-----CGCATCTGTGCTGCTGAGCG-----CCCTC 508
DB 1900 CGGTACGATGACGGTGTCTAGAGCTCATCTTATGAC--CGGAACCAATCGACGTCTTC 1958
QY 509 CCCAGGAGCGGCG-----GAC-----CC-----CAGCCATC--CAGG-- 540
DB 1959 CACCAAGTTGCTATGTAAGCGCGTTGATGTTCTCTCGTCTCTCTATCTCTGTTA 2018
QY 541 ---GGCG--AAGGGAATTAGTCTGTG--GGTCTCCCC-----AAGCGGCTC 586
DB 2019 TACTGGCGTTGATCGAGTACG-----CCTGTGTTCCCTACTCAAAAAGAGACAGGATC 2075
QY 587 GCGGATCTGAGGAGAACAAAGCCGATCGGCGGCC--ACTGCGCC-----CT 632
DB 2076 GTCGGA--GAAGAGAGAAAGACGAGCATAAACCTGCTCCGCGGACACCTGATATTCT 2133
QY 633 TAATCTGATCCAGCTCGGGCTGAGGCTGAGGCACTGGCGGAGGAGGCGCG--TCCTCT 690
DB 2134 TCAGAGCTCC-GCCT--TGCCGA-----ATGCATGCAACGCG--GGCTCAACCTCG 2182
QY 691 CTGCACACC-----TACT-----AGTCAC-----CAGAGACTT- 718
DB 2183 ATCATCGCGCTCATCAACAGCAGTCGAATTCCTGTGTCAGTCACATTCAGTTCATC 2242
QY 719 -TAGGGGT--GGGATTCCACTCGT-----GTGTT--TCTATTTTTT- 756
DB 2243 GTCAGCGGTGCGGGTTTCTCTGTTTTCATCTGTTTCAACTCTCTCTGCGGTGATT 2302
QY 757 -----AAACGACAGCAATT-----TTAAAAATGGTCAAG-----TTTGGT 791

Query Match	41.8%	Score 2412	DB 4	Length 2508
Best Local Similarity	39.5%	Pred. No. 1.9e-101		
Matches 578	Conservative	0	Mismatches 316	Indels 570
Gaps	81			
2	GGCC	-----CATCTGTTTC-----AGCC-----AGT- 24		
1074	GGTCTGGGANGGTTTCTTAATTTTGTTCGTAAATGGTATTCGCCCTCGTTCGAGTA 1133			
25	CGCCA-AGAAAT	-----CATGAA-AGT'CG-----CCAGTGGCAGCA-----C 58		
1134	CGCCATAGTATCTCATATGAATAAACGACTGGTCTCCGACGGGAAAAACGAGAAAAGC 1193			
59	CGCC-ACCGCGCGCGGGGCC-----CCAG-----CTGCGCG---CTGAAAGCCGGCAA 103			
1194	CGCGAACACAGCAGCGAAACGAGATGCCATGTTCAACGCGAGCCGGAAGCCGCCAA 1253			
104	GA-----CAGCGAGCGGGTG-----CGGCGCA---GGTGGT 130			
1254	TAATAATAATCATTGATGGAGATCCACAGAAAAATGTGATTCGCGAGCATTCCAATGAT 1313			
131	GGCGTGTG-----TGCTGAG-CAGAGC-----GTGGCCATC-----TCGGGCT 168			
1314	GCAACATCCACGTCTTGTCCAGACGGCGCACATACGCTATGGCCGGCTCCATTCGCGCG 1373			
169	GCCG-----GGCG-GCCGGGGCGGCTGCCCTGC-----CCTGC-----TGACGAGC 210			
1374	GCCGAAAAAGGTTTCCAGACATG-CTGCCAACGATGGACGCCGTGCAAAATCGATAAGC 1432			
211	--AGCAGGTA-----AACGTGCTGCTCTA-----CGACATGAACGGCTG----- 247			
1433	TTAGCCGATACGGTTTCCCAATGTCTTTCTCTATCTTCAATATAGTCTACTGGTTGTATA 1492			
248	-----TTACTTCAGGC--TCRAAGGAGCTGGTGCCACCCT--GCCCCAGA 288			
1493	TGAATATCTAAGCTTTAACTCGTCGCAACAGATCCAGGAGACGACAGATGGGACGAGA 1552			
289	ACGCAAGGTGAGCAAGGTGGGATTTCTCCAGCAGCTCATCGACTACATCAGG-GACCTT 347			
1553	TCAC-----TGAATGGTATTTCAGCGC--CGAAATCGAGTACAAATGGTGTACGT 1601			
348	CAGTTGGAGCTGAACCTCGGAATCCGAGTTGGGACCCCGGGGGCCGAGGGCTGCGG--- 404			
1602	C-GAAGGACCGAATGTTTCGACAGCGGTCAAGCGCGACGCGAACATCGAACTGTGAGT 1660			
405	-----GTC-----CGGGCTCCGCTCAGCAC-----CC--- 426			
1661	TATAAATTCACATAAATCTGCCAANAAGGACACTTCCAGCAGCTTCATCGGGAGCTAC 1720			
427	-----TCACGG-----CGAATCAGCG----- 444			
1721	TCGCTGCTACGGGTTAGTTTCATATTTGATCCGGA-CAGCGGCTTCTACTTCTTCAAT 1779			
445	-----CCCTGACGGC----- 457			
1780	ATTTTTCCTCGCAGCTCGTGTAGTTTATCATGGATCTCATCTGGATCAATCGTGA 1839			
458	-----GGCGGCAT-GCGTTCC-----TGCG--GACGAT-----CGCATTTGT 492			
1840	CTCGGCGCTCTCGCAACCCCTAATCGGTACGATACGCGTGCTCACTGAGACTCATCTTAT 1899			
493	GTGCTGTAAGCG-----CCTCCCGCAGGACCGCG-----GAC-----CC 528			
1900	GAC-CGGAAACCAATCGACGCTCTTCCACGATGTGCTATGTAAGCGGTTGATGTATTC 1958			
529	-----CAGCCATC--CAGG-----GGGC-AAAGAGAAATTACGTGCTCTGTG-GGTCT 571			
1959	TCGGTTTCTGCTATCTTCTGGTTATACTTCTGGGTTGATCGAGTAGC---CCTGTGTTGCT 2015			
572	CCCC-----AACGGCGCTCGCCGATCTCGAGGAGACAGACCCATCGGCGGC 622			
2016	ACTCAAAAAGAGAAACGAGGATCGTCCGA--GAGAGAGAGAGACGAGACATAAAC 2073			

445	-----CCTGACGGCC-----	-----GA 457
1780	ATTTTTCCTCGACGCTCGTAGTWTTTATCANGATCTCAATCTGGATCAATCGTGA	1839
458	-----GGCGGCAT--GCGTTCC-----TGCG--GACGAT-----CGCATCTTGT 482	
1840	CTCGCGCCTTCGCGAACCCTAATCGGTACGATGACGGTGTCTCACTGAGACTCATCTTAT	1899
493	GTGCTGAGCG-----CCTCCCCAGGACGGCG-----GAC-----CC 528	
1900	GAC--CGGAACCAATCGACGTCTTCCACAGTTGCTATGTAAAGCGGTGATGTATTC	1958
529	-----CAGCCATC--CAGG-----GGCG--AAGAGGAATTACGTCTCTGTG--GGTCT 571	
1959	TCGGTTTCTGCTATCTTCTGGTTATCTTGGCTTACTGGCGTTGATCGAGTACG---CCTGTGTTGCGCT 2015	
572	CCCC-----AACGGCCTCGCCGATCTGAGGGAGAACAGACCCGATCGGCGGCC 622	
2016	ACTCAAAAAGAGAAACGAGGATCTCGCA--GAAGAGAGAGAGACCGAGCATAAACC	2073

156 -GCCATCT-CGGGC-----TGCC-----GGGGGG-----CGGGG-----GGGCT 190
417 CACCATCTACACGCTGGATGCTGGCTCTTTGGGGCCCTCGTCTCAAGGCGCGTCACT 476
191 GC-----CTGCCCTGC--TGGACGACAGCAGGTAAAGT--GCTGC--TCTAC--GACA 237
477 GCTCATCTACCTACCATGATGCG--CAGCAGCTTTACGCTGCTGCTCTCTCCGTTGACA 535
238 TGAAC--GGC--TGTTACTACGC-----CTCAAGAGCTGTGTGCCACCTG--CCCCA 286
536 GGTACCTGGCGTGGGACCGCTGCTGCTGCGGCGCTGCGCACGCGCGGTAAACGCC 595
287 GAACCGCA-----AGGTGACCAAGGTGAG-----ATTCTC--CAGCAGTCACTGAC 332
596 GCGCCGACGTGGGGCTGTGTGCTGCTGCGGCGCTTCTTCGGC--GCCCTACTCTAGC 654
333 TACATCAGGACCTTCAGTTG-----GAGCT-GAATCTCGGAAT--CCGAAGTTGGA----- 381
655 TACTAC--GGCAGCTGCTGACGCGCTGAGGCTCTGCTGCGCGC--TGGAGGAC 711
382 -----CCCCCGG-----GGCGAG-----GCTGCGG--TCCGGGCTCGCTCAGC-- 422
712 GCGCGCGCGCGCGCTGACGCTGCGCTGCTGCTGCGGCTACTGCTGCGCGTGGCT 771
423 -----ACCTC--CAAC--GGCAGATCAGCGC--CCT---GACGCGCGAGG--CGGCA 464
772 GTGGTGAAGCTGACCTACGGGCGCAAGCTGGCTTCTGTGGGCGCGCTGGTCCGCG 831
465 TGCGTCTCTCGGACGA--TGCATCTTGTGCTGCTGAAGCGC-----CTCCCC----- 511
832 GCGCGCGCGCGCGCGAGGCGCGGAGGCGGACGCGCGCGCGCGCGCGCTATGCTG 891
512 -CAGGACCGCGGACCCC--AGCATCCAGGGGCAAGAGGAATTAAGTGTCT---CTGT 565
892 GCGGTGGCC--GCGCTCTACGCGCTCTGCTGGGTCC-----GACACGCGCTATCTGT 946
566 GGGTCT-----CCCCCAACGC-----GCTC-- 586
947 GCTTCTGTAGCGCGCTTCCGCTTACGCGCGCACTACGCTGCGCGCTGGCTCAC 1006
587 ---GCGGATCTAGGGAGAACAG-----ACGATCG-----GCG---GCCACTGCGC 629
1007 ACTGCTGGCTTACGCCAATCTGCTGCTCAACCGCTGCTTACGCGCTGCTGCTC--GCC 1065
630 CTTTAAC-----TGATC--CAGCCTGGGC--TGAGGCTG--AGGCACCTGGCAGAG-A 678
1066 CACTTCGCGCGCTTCCGCGCTTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1125
679 GGGCGCTCTCTGACACCTACTAGTACACAGACTTTAG--GCGT-----GCGT 726
1126 GCGCGCGCTTGGCTGCGCTGCGC-----CCGCGCTCTGCGGCGCGCGCGCGCGCG 1179
727 -GGGATTC--ACTGCTGTGTTCTATTATTTTGAAGCAGACA-----TTTTTAA 774
1180 GGAGACGCGCGCTACGGGAGGCTGCTGCTG--GTGGGCGCGCGCGCGCGCGCGCG 1238
775 AAATGGTCACTTTGGTCTTCTCAGATTTCTGAGGAAATGCTTTGTATTGTATTAT 834
1239 GGAGGACCCCTCCACGCGCGAGAGGCTGCGCGAGGAC-----GGAATAAACCTCTC 1291
835 --AATGATCAGGACTGAGATAT--TGTTTACATATGTTCT-----GTGGG----- 879
1292 GCGCTGAGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1351
880 -----CTGTTTTTTTGTATTAAACAAATAATTTAGATGTGAAAAAAA 925
1352 CGGGCAGGGATGGGCAATGCCACGAGCTCTCTGAGGGCGGTGAGTGA 1402

RESULT 3
US-08-977-767-3
; Sequence 3, Application US/08977767
; Patent No. 5972684

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Yue, Henry
APPLICANT: Greenwald, Sara
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ For Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,767
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0423 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1532042
US-08-977-767-3

Query Match 42.2%; Score 2431.5; DB 2; Length 1345;
Best Local Similarity 42.2%; Pred. No. 1.6e-102;
Matches 564; Conservative 0; Mismatches 301; Indels 470; Gaps 89;
QY 1 GGGGCCCCATTCTGTTTTCAGCCAGTCGCCAAGAATC-----ATGAAAGTGC--CAG-- 49
Db 72 GGGGTCC--CAG-ATC--CCCGCTGCTCTGTAAGTGGGAACTGGGACTCCCTGCAGGA 125
QY 50 -----TGGCAGCACCG-----CCACC----- 65
Db 126 GGAGCGGAAGGCTGAAGTCTCTGGAGGGATTAGGGTTCCCAAGAGGGGGAGAGCTG 185
QY 66 -----GCC-GCCGCGGCG-----CCGAGCTGC-----GCGCTGAAGG 96
Db 186 AGAGAGGCG 244
QY 97 C--CGCGCAAGACAGCAGCGCGGTGGCGGAGAGGTGGTGGCT--GCTGTCT--GAGCAGA 150
Db 245 CTGCACCTCTGTCAGCGCGCTCGGCTGT--TGGTACTCTGGGCGCGCGCGCGCGCGCG 302
QY 151 GGGT-----GGCCATCTC-----GCGCTGCGGCGG-----CGCCGG 181
Db 303 ACGTCACATCGGACCCCGCACCTTGACAGGACTGGTGGAGTACAGGATATCTCAGG 362
QY 182 GG-----GCGCGCTG--CCTGCGCTGCTGAGCAGCAGGTAACCTG--CTG---C 228
Db 363 GAAACTTCGTGCGCAGGACCT--CCCTTCTG---GGGC--CTGGTGAACGACGCTGGAGTC 416

Db 398 GATCCACTCCTA TGTGACCGGGTCTGACTCCAAACGAGCAGACGAAAGATGAGCTTCAAGG 457
Qy 217 TAAAGCTG---CTGC-----TCTAC---GACATGACGGCTGTCTTACTCAGCC--- 258
Db 458 AGATCAAGAGCTGCTGGAATGCTCAACGCTGACATGAACGACATGTAGCGCTACTCTCC 517
Qy 259 ---TCAAGAGCTGGTCCCAAC-CTGCCCCAGAACCG-CAAG-GTGAGC-AAGGTGGAG 311
Db 518 TCTTCAAGAGT--GTGACCACTCCAAAC--GACCGCTAGAGGGGGCTGAGATCGAG 573
Qy 312 ATCTCCAGC-ACGTC-----ATCGACTACATCAGG---GACCTTCAGTTGGAGCTGA 360
Db 574 GAGTTCCTGCGCGGCTGCTGAACGGCGGAGCTGAGGAGATCTTCAT-----CAGT 628
Qy 361 ACTCGG-----AATCG-AAGTTGGACCCCGGG---GCCAGGG-CTGCCGGTCC---- 408
Db 629 ACTCGGGAGAGCGCGTGTGTAGTGCCTGTAGCTGTGGATTTCTTGGAGGACAG 688
Qy 409 -----GGGCTCGCTCAG---CACCTCAACGGCGAGATCAG-----CGCCC 447
Db 689 GCGAGGAGGGCCACACTGCGCCCGCGCCAGCAGCTCATTCAGACCTATGAGCTCAAG 748
Qy 448 TGACGGCGAGCGGCTGCTTCTGGGAC-----GATCGATCTTG-----TG 493
Db 749 AGACAGCC-AAGCAGCATGAG---CTGATGACACTGGATGCTTCTGATGTACCTGTGTG 804
Qy 494 TCGCTGAAG-----CGCCTCCC-----CCAG-----GGACCGG 521
Db 805 TCGCGGAGGGGACTGCGCTTGGACAAACCCACACGCTGTGTGTTCCAGGACATGAACAG 864
Qy 522 C-----GACCCCGC--CATC-----C-----C 537
Db 865 CCGCTTGCCCACTATTCTATCTCTCTCCACACACCTATCTGACTGCTCCCACTC 924
Qy 538 AGGGGCAAGAGGATTAACGTG---CTCTGT---GGGTCT---CCC-----CCAGCGCG- 582
Db 925 GGGGGCCC-AGCAGCAGGAGCTATGTAGGGCTTTGGCCAGGGATGCGCTGCTGT 983
Qy 583 -----CCT 585
Db 984 GAGCTGAGCTGCTGGAGGGGCCAGGAGGGAGCCCGTCACTATCATGCGCCATACCT 1043
Qy 586 CGCC-----GGATCT-----GAGGAGAACAGACCGATCGCGGCCCA- 623
Db 1044 CACCTCCAGATTCTTTCGGGAGCTGGCCCAAG-CGGTGC-GGACCATGCTTCAAG 1101
Qy 624 CTGCCCTTTA-ACTG-CATCAG--CCTGGG-----CTG-AGGCTGAGGCA-CTGGCG 672
Db 1102 CTGTCCCTTACCTCTGTCTATCTATCTGAGAACCACTGCGGGCTGGAGCAGCAGGC- 1160
Qy 673 AGGAGAGGGCGCTCT-CTCTGCAC-----ACCTACTAG----- 705
Db 1161 -TGCCATGGCCGCGACCTCTGACCATCTGCGGGGACATGCTGTGACACAGGGCGTGG 1219
Qy 706 -----TCACGAGAG--ACTTTAGG-----CGT-- 726
Db 1220 ACTCCCCAAATCCGAGAGAGTGCAATCCCGAGAGAGCTGAAGGGCGGGTCTCTGTGA 1279
Qy 727 -GGGATT-----CCTGTGT-----TTCTAT----- 750
Db 1280 AGGGAAGAGAGGTGCGCCGCTCTCGGAGCGAGGATGCGCGGCTCTGTGATCGGGAGG 1339
Qy 751 -----TTTTTGAAGAGAGACATTTTAAAAATGCT-----CA-----CGTT 787
Db 1340 AGGGGAGAGAGATGACAGGAGGAGAGAGAGGAGGTGGAGGCTGACGCGAGAGCGGCG 1399
Qy 788 TGGTCTCTCTCAGATTTCTGAGGAAT---TGCTTTGTATGTATATTAATGATCACC 844
Db 1400 TG--GCCAAGCAGATCTCCCCGAGCTGTGCGCCCTG-GCTGTGTACTG-CACGCCACCC 1456
Qy 845 GACTGAGATATTGTTT-----TACATAGTTCTG-TGGGGCTG-TTTTTT 889
Db 1457 GCCTCGGACCTGTGACCTGCGCCCGCCCAAGCGCCCAACCTGCCAGGTGAGCTCCCTCA 1516

Qy 890 GTTA---TTAAACAATAATTAGATGGTGAAAAAAA 926
Db 1517 GCGAGCGCAAGCGCAAAACTCATTCGGAGGCGGAA 1556

RESULT 2

US-08-900-230-3
Sequence 3, Application US/08900230
Patent No. 6329197
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.
TITLE OF INVENTION: DNA ENCODING GALANIN GALR3 RECEPTORS AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of The Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 11036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,230
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52241-C/JPW/ADM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO
HYPOTHEetical: NO
ANTI-SENSE: NO
US-08-900-230-3

Query Match 44.4%; Score 2559; DB 4; Length 1417;
Best Local Similarity 43.0%; Pred. No. 3,1e-108;
Matches 564; Conservative 0; Mismatches 331; Indels 416; Gaps 86;

Qy 1 GGGGC-----CCA--TTCTGTTC---AGCCAGTC----- 25
Db 122 GGGGCTTCTGCTGAGATTCCAGCTTCTTCCAGTGGCCGCTGATGGGAGATGG 181
Qy 26 -----GCCAAGAA-----TCATGAAGTC-----GCCAGTGGCAGCAC- 59
Db 182 CTGATGCCAGAACATTTCACTGGACAGGCCAGGAGTGTGGGGCC-GTGGCAGTGCCT 240
Qy 60 -----GCC-----ACCGC---CGCGCGGGCC----- 78
Db 241 GTGGTCTTTCCTTAATCTTCTGCTGGGCAAGTGGGCAATGGGCTGGTGTGGCAGTG 300
Qy 79 --CCAGCTGC---GCGCTGAAGGCC--GGCAAGA-----CAGC-----GAGCGGTGG- 119
Db 301 CTCTGCAAGCTGGC--CCGAGTCCCTGGCAGGAGCTGGCAGCACCCAGGACCTGTTTCA 359
Qy 120 ----GGCGAGGTGTGGCTG---TCTG-----TCTGAGCAGAGGCTG----- 155
Db 360 CTTCAACCTGGCGTG-GCTGACCTCTGTTCACTCTGTTC--TGCGTCCCTTCCAGGC 416

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:17:00 ; Search time 23 seconds
(without alignments)
2078.506 Million cell updates/sec

Title: X77956
Perfect score: 5766
Sequence: 1 GGGGCCCATCTGTTTCAGC.....TTAGAGTGTAAGAAAAA 926

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 11 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A COMB pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2596	45.0	2211	4	US-09-738-884-1
2	2559	44.4	1417	4	US-08-900-330-3
3	2431.5	42.2	1345	2	US-08-977-767-3
4	2412	41.8	2508	4	US-09-627-650B-7
5	2412	41.8	2508	4	US-09-436-063C-7
6	2397	41.6	2544	4	US-09-627-650B-3
7	2397	41.6	2544	4	US-09-436-063C-3
8	2385	41.4	2088	4	US-09-548-372D-13
9	2385	41.4	2088	4	US-09-548-372D-13
10	2385	41.4	2088	4	US-09-551-853D-13
11	2383.5	41.3	2601	4	US-09-627-650B-9
12	2383.5	41.3	2601	4	US-09-436-063C-9
13	2337.5	40.5	1400	4	US-08-630-915A-37
14	2327	40.4	1652	4	US-09-627-650B-1
15	2327	40.4	1652	4	US-09-436-063C-1
16	2314	40.1	1497	4	US-09-060-854B-2
17	2267.5	39.3	1917	4	US-09-627-650B-5
18	2267.5	39.3	1917	4	US-09-436-063C-5
19	2240	38.8	1128	4	US-09-627-650B-11
20	2240	38.8	1128	4	US-09-436-063C-11
21	1965	34.1	1388	4	US-09-463-048A-6
22	1772.5	30.7	908	4	US-08-714-741-44
23	1479.5	25.7	717	4	US-09-644-460-37
24	1355.5	23.5	605	4	US-09-976-594-616
25	1311	22.7	801	1	US-07-906-349A-6
26	1265.5	22.0	1184	4	US-10-153-064-8S
27	1110	19.3	1332	1	US-07-609-716-41

28 1110 19.3 1332 3 US-08-475-411A-41 Sequence 41, Appl
29 1110 19.3 1332 3 US-08-478-029A-41 Sequence 41, Appl
30 942 16.3 1136 4 US-08-806-029-9 Sequence 9, Appl
31 942 16.3 1177 1 US-07-609-716-31 Sequence 31, Appl
32 942 16.3 1177 1 US-08-175-155-29 Sequence 29, Appl
33 942 16.3 1177 1 US-08-477-509B-64 Sequence 34, Appl
34 942 16.3 1177 2 US-08-707-237A-35 Sequence 35, Appl
35 942 16.3 1177 3 US-08-482-085B-64 Sequence 64, Appl
36 942 16.3 1177 3 US-08-475-411A-31 Sequence 31, Appl
37 942 16.3 1177 3 US-08-478-029A-31 Sequence 31, Appl
38 942 16.3 1177 4 US-09-444-791A-64 Sequence 64, Appl
39 905.5 15.7 1059 1 US-08-175-155-48 Sequence 48, Appl
40 905.5 15.7 1059 2 US-08-707-237A-54 Sequence 54, Appl
41 905.5 15.7 1059 2 US-08-806-029-10 Sequence 10, Appl
42 905.5 15.7 1101 1 US-08-477-509B-83 Sequence 83, Appl
43 905.5 15.7 1101 3 US-08-482-085B-83 Sequence 83, Appl
44 905.5 15.7 1101 4 US-09-444-791A-83 Sequence 83, Appl
45 902.5 15.7 1038 1 US-07-609-716-36 Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-738-884-1
; Sequence 1, Application US/09738884
; Patent No. 6391606
; GENERAL INFORMATION:

; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000849
; CURRENT APPLICATION NUMBER: US/09/738,884
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2211
; TYPE: PRT
; ORGANISM: Human
US-09-738-884-1

Query Match 45.0%; Score 2596; DB 4; Length 2211;
Best Local Similarity 38.8%; Pred. No. 9.3e-110;
Matches 597; Conservative 0; Mismatches 305; Indels 638; Gaps 89;

Qy 1 GGGGCCCATCTGTTTCAGCAG-----TCGC-----CAAGATC----- 35
Db 41 GGGGCTC---CCGCTCCGCAAGATCCGCTCGGCACGTGGCACAAGAGCGGTGTACC 97
Qy 36 -----ATGAAAGTC-GCCAGTG-----GCAGCAC-----CGCCACCGC 67
Db 98 GGCTGCAGGAGGACGGCTGAGCGTGTGTTCCAGCGGGCATCCCGGTGGCGCATCGC 157
Qy 68 CGC-----CGGGGGC-CCAG----- 82
Db 158 AGCAGATCTTCTTCGTGTCAGCATCGAGGGGTTCGCGAGGGCCACCAGTCCGAGGGCC 217
Qy 83 -----CTGCGCGCTG-----AAGGCC 98
Db 218 TCGCGGCTTCGGGGGTGCTTCGCGCCAGCGGCTCCCTCACCATCGCTTCAAGGGCC 277
Qy 99 G--GCAGAG--CAG---CGAGCGGTG---CGGGCGAGGT---GGTGGCT 135
Db 278 GCGCAGAGAACCTTGAGACCTTGGCGGCCCCACGCTGAGGAGCGAGCGTGGTGGCG 337
Qy 136 GTCTGTGTAGCAGACGCTG-----GCCATCTCGCTGC-----CG 172
Db 338 GTCTGACCAAGCTCCCGCGCGCTTGAGCGCCATGAGCAGCGCGAGCGGTAGACACT 397
Qy 173 GG-----CGCGCGGGG--CGCCTGCTCCCTGCTGGAGCGAGCAGC--AGG 216

QY	632	TTAACTGCATC--CAGCTGGG--CTGAGGCTGAGGCACTGGC-----	671
Db	980	TACCTTGC-TCTGCTGATGGGCCCCCATGGCTTTGG--CTGCCACTGAGGGTAGGGTG	1036
QY	672	--GAGGAGAGGGCGCTCTCTCTGCAACCTACTAGTC--ACCAGAGACTTTAGGGGGTG	727
Db	1037	TGGAGGTGTGGAGGCCCC--CTGAGGAGCTGC--GGCGGCCAGGTAC---GAAGCTG	1087
QY	728	GGATTCCACTCG-TGTG-----TTTCTATTTTGAARAGCAGACATTTTAAATGAT	781
Db	1088	CAACTCTGGCGCAGTGGGCGAGATCTCAT-----CAGCC-----CCAGGC	1128
QY	782	CAGTTTGGTGTCTTCTCAGATTTCTGAGG-----AAATTGCTTTGTATTGTA	828
Db	1129	TGCAGGTGAGGCTTCAGGGGATCTGGGGCCCACTGCCCTCGCTGCTTGGCCCTCA	1188
QY	829	T-ATTACATGATC-----ACG-----ACTGAG-----AATATTGTTT--TA	863
Db	1189	TCCTTCCTCTGTTCTTCTGCGCGGCAACACAGCACTGGGGCTCACCTCTTGGTTGATC	1248
QY	864	CAATAGTTCGTGGGGCTGT-TTTTTTGTATTAAACAATAATTTAGATGGTGAAAA	921
Db	1249	CTCTTGTACTG-GGAGAGGTGCCTTTTGT--ATCCCCAATTA-----AAGGTAGAAA	1298

Search completed: May 7, 2004, 15:17:53
Job time : 85 secs

```

PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 25-JUN-1998; 98US-0090697P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095988P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98US-01019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100949P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 25-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.

PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-00168978.

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred.No. 3.5e-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

QY 1 GGCGCCCATTCGTTTTCAGC-----CAGTCGCCAAG---AATCATGAAGTGGCCAGTG 51
DB 40 GTGGCAGCGCGGTAGCAGCAATGGACTTCTCTGGGGAACCCGTTACGCTCTCCAGTG 99
QY 52 G--CAGCAC--CGCCACCGCGCGCGGCG--CCAGCTGGCGCTGA--AGGCC----- 98
DB 100 GGACGCGCATCGAGAAAGCCACAGATGCTCCCTCAGAGGAGACTGGGCGCTTAAC 159
QY 99 --GG-----CAAGACAG-----CGAGCG 114
DB 160 ATGAGATCTGGACATCATCAAGAGACGGAGGAAGTCCCAAGATGCCCTCCGAGCA 219
QY 115 GT-----CGGGG-----CGAGTGGTGC--CTGT--CTGTCT 142
DB 220 GTAAAGAGAGATCGTGGGAATAAGAACTTCCACGAGGTGATGCTGCTCTACAGTC 279
QY 143 TGAGCA-----GAGCGTGGCCATCTCGCGT-----GCCGGGCG 176
DB 280 TTAGAAACCTGTGTCAAGAACTCGCGGCACCGCTTCCACGTGTGTGGCCAGCAGGAC 339
QY 177 GCCGGGCGCG-----CCTGCTGCCCTGTGGACGAGCAGAGTAAACG 222
DB 340 TTCTGGAGAGTGTGCTGTGGAGACCATCTGCGC--AAGAACAAACCCACCATCG 397
QY 223 TGC-----TGCTCTACGACAT--GAAC--GGCT-----GTTACTCA--CGCC-- 258
DB 398 TGCATCAAAAGTGTCAACCTCATCCAGTCTGGGCTGACGCTTCCGAGCTGCCCG 457
QY 259 -TC---AAGGAGTGTGGCCCA-----CCCTGC-----CCC--AGAACCGC 293
DB 458 ATCTGACAGTG--TGCTCAACATCTATGAGGACCTCGGAGAAAGCGCTGGAGTTCGCC 516
QY 294 AAG-----TGAGCAAGTGGAGATCTCCA-----GCAAGTCATGACT-- 333
DB 517 ATGACTGACCTGGACATGTGTCAACCATCCACACCCAGAGGACCGTGTCAACTCAG 576
QY 334 ---ACATCAGGAC---CTTCAGTTGGAGTGAATCGGAATCGGAATGGAGCCCG 386
DB 577 AGACACAAATCAGGACAGGATTTCTGTGGCAGTGTG--ACTC--CAGCCAGCAAGAGACTCT- 632
QY 387 GGGGGCC--GAGGCTCGCGTCCGGCTCGGCTCAGCAC--CCTCAA--CGG---CGAGATC 440
DB 633 ---GGCCAGCATCTGCCCTCTG-----CCGCGCCCGCCCATCTCTCCGCTGACACGCC 686
QY 441 AGCGCCCTGACGGCCGAGGCGGCGATCGTTCTCTGCG-----GACGATCG----- 485
DB 687 ATAGCACCAAC--CCCGCAACAGATTGGGAAGCTGCGCAGTGGAGTGGAGTGGTGG 745
QY 486 -ATCTTG-----TGTC-----GCTGA--AGC-----GCCTCCCGCCAG--GGACCGG 521
DB 746 GAAAGTGGAGGTGATGTCCGAGATGTCTGAGGAGTGGTGCC--CACCCAGCGCCAGCCCG 804
QY 522 CGGACC-----CC--AGCCATCCAG-----GGGGCA 545
DB 805 CAGACCTGGAGCTGTGACGAGGCTCAACCGCAGCTGCCGAGCCATGACGACGCGGTCC 864
QY 546 AGAGGAATTAGTGTCTGTGGGTC--TCCCCAACGCGCC--TC---GCCGATCTGAG 598
DB 965 TGAGTGATACCTCTGCTC--CGGGCCCATGCCCCAAGAGGCGCTTACAGAGCCACACTGCC 922
QY 599 GGAGAACAAAGACCGATC---GGC--GGCCACTG-----CG-----CCC 631
DB 923 AG---TCGAGGCTGGCTGGAGGCTGGCCACAGTGGAAATTCCTGCCGAGCTATTGTC 979

```

```
QY 672 --GAGGAGAGGGCGCTCTCTCTGCACACCTACTAGTC--ACCAGAGACTTTAGGGGTG 727
Db 1037 TGGAGTGTGGAGGCCCC--CTGAGGAGCTGC--GGCGGCCAGGTAC--GAAGCTG 1087
QY 728 GGAATCCATCG-TGTG-----TTTCTATTTTGAAGAAGCAGACATTTAAATAATGGT 781
Db 1088 CAACTCTGCGCGAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC 1128
QY 782 CACGTTTGTGTTCTCTCAGATTCTGAGG-----AAATTGCTTTGTATTGTA 828
Db 1129 TCGAGTGAAGCTTCAGGGAGTCTGGGCCCCACCTGCCCTCGCTTGCCTCCCA 1188
QY 829 T-ATTACATGATC-----ACCG-----ACTGAG-----AATATTGTTT--TA 863
Db 1189 TCCTTCCTCTGTTCTCTCTGCGCGGCCACACAGCACTGGGGCTCACCTCTTGGTGTATC 1248
QY 864 CAAATAGTCTGTGGGCTGT-TTTTGTGTTTAAACAATTAATTAAGATGTTGAAAAA 921
Db 1249 CTTCTTACTTG-GGAGAGTGGCTTTTGT--ATCCCCAATTA-----AAGGTAGAAAA 1298

RESULT 15
ABU88959
ID ABU88959 standard; protein; 1300 AA.
XX
AC ABU88959;
XX
DT 09-JUL-2003 (first entry)
XX
DE Novel human secreted and transmembrane PRO protein #2.
XX
KW Human; gene therapy; chondrocyte stimulation; TNF-alpha release;
KW chondrocyte proliferation; chondrocyte differentiation; tumour detection;
KW tissue typing.
XX
OS Homo sapiens.
XX
PN US2003036133-A1.
XX
PD 20-FEB-2003.
XX
PF 27-JUN-2002; 2002US-00184630.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-00631121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090439P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
```

```
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095988P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 18-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 01-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 02-SEP-1998; 98US-0098843P.
PR 09-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 10-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 17-SEP-1998; 98MO-US019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100919P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100849P.
PR 18-SEP-1998; 98US-0101014P.
PR 23-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 25-SEP-1998; 98US-0101922P.
PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102303P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102985P.
PR 06-OCT-1998; 98US-0103258P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-00168978.
PR 07-OCT-1998; 98US-0103395P.

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred. No. 3 5e-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

Qy 1 GGGGCCCATCTCTTTTCAGC-----CAGTCGCCAAG---AATCATGAAGTGGCGAGTG 51
Db |||||
Qy 40 GTGGCAGCGCGGTAGCAATGGACTTTCTCTGGGAACCCGTTAGCTCTCCAGTG 99
Db |||||
Qy 52 G---CAGCAC--CGCCACCGCGCGCGGC--CCACAGTGCAGCTGA--AGGCC----- 98
Db |||||
Qy 100 GGACAGGGCATCGAGAAAGCCACAGATGCTCCTCCAGAGGAGGAGGAGTGGGCGCTCAAC 159
Db |||||
Qy 99 --GG-----CAAGACAG-----CGAGG 114
Db |||||
Qy 160 ATGGAGATCTGCACATCATCAACGAGACGGAGAGTCCCAAGATGCCCTCCGAGCA 219
Db |||||
Qy 115 GT-----CGCGG-----CGAGGTGGTGC--GCTGT--CTGTC 142
Db |||||
Qy 220 GTAAAGAGAGATCGTGGGAATAGACTTCCACGAGGTGATGCTGGCTCTCACATC 279
Db |||||
Qy 143 TGAGCA-----GAGCGTGCATCTCGCGCT-----GCCGGGC 176
Db |||||
Qy 280 TTAGAAACCTGTCTAAGAACTGCGGGCACCGCTTCCAGGTGCTGGTCCAGCCAGGAC 339
Db |||||
Qy 177 GCCGGGCGCG-----CCTGCCTGCCCTGCTGGACGAGCAGCAGGTAAACG 222
Db |||||
Qy 340 TTCGTGAGAGTGTGCTGTGAGGACCATCTCGCCC--AAGAAACAACCCACCCCATCG 397
Db |||||
Qy 223 TGC-----TGCTTAGACAT---GAAC--GGCT-----GTTACTCA---CGCC-- 258
Db |||||
Qy 398 TGCATGACAAAGTGTCTCAACCTCATCCAGTCTGGGCTGAOCGTTCCGAGCTCGCCCG 457
Db |||||
Qy 259 -TC---AAGGAGTGTGTGCCCA-----CCCTGC-----CCC--AGAACCGC 293
Db |||||
Qy 458 ATCTGACAGTG--TGTACCATCTATGAGGACCTGCGGAGGAAAGGCTGGAGTTCGCC 516
Db |||||
Qy 294 AAGG-----TGAGCAGGTGGAGATTCTCCA-----GACGTCATCGACT--- 333
Db |||||
Qy 517 ATGACTGAOCTGGACATGCTGTACCCATCCACACCCAGAGGACCGTGTTCACATCAG 576
Db |||||
Qy 334 ---ACATCAGGGAC---CTTCAGTTGGAGCTGAACTCGGAATCCGAAGTTGGGACCCCC 386
Db |||||
Qy 577 AGACACATCAGCAGCAGGATTCTGTGGCACTG-ATC--CAGCCAGCAGAGGACTCT- 632
Db |||||
Qy 387 GGGGGC--GAGGGTCCCGTCCGGCTCCGCTCAGCAG--CCTCAA CGG-----CGAGATC 440
Db |||||
Qy 633 ---GGCCAGCATGCTGCCCTCTG---CCGCCGCCGCCCACTACTCTCGGTGACACGCC 686
Db |||||
Qy 441 AGGCGCTGACGGCCGAGCGCGCATCGTTCCTGCG-----GACGATCGC----- 485
Db |||||
Qy 687 ATAGACCAAC-CCCGAAACAGATTGGGAAGCTGCGCAGTGGAGTGGTGGTGG 745
Db |||||
Qy 486 -ATCTTG-----TGTC-----GCTGA---AGC---GCTCCCGCAG---GGACCGG 521
Db |||||
Qy 746 GAACGTGAGGGTGTATCTCGAGATGCTGACGAGGTGGTGCC-CACCCAGGCCGAGCCCG 804
Db |||||
Qy 522 CGGACC-----CC-AGCCATCCAG-----GGGGCA 545
Db |||||
Qy 805 CAGACCTGGAGCTGTGACGAGCTCAACCGCAOCTGCGAGGCCATGACAGCAGCGGTCC 864
Db |||||
Qy 546 AGAGGATTACGTGCTCTGTGGTC--TCCGCCAAGCGCC--TC---GCCGATCTGAG 598
Db |||||
Qy 865 TGAGTGATACCTGTCTC--CGGGCCCATGCCCAGGAGCCCTTCAGAGCCACACTGCC 922
Db |||||
Qy 599 GGAGAAACAAGACCGATC---GGC--GGCCACTG-----CG-----CCC 631
Db |||||
Qy 923 AG--TCGAGGCTGGCTGGAGGCTGGCCACAGTGGAAATCTGCGGAGCCTATTGTGCC 979
Db |||||
Qy 632 TTAACTGCATC--CAGCCTGGGG--CTGAGGCTGAGGCACTGGC----- 671
Db |||||
Qy 980 TACCTGC-TCTGCTCATGGGCCCCCATGGCTTTGG--CTGGCCACTGAGGGTAGGGTG 1036
Db |||||
```

Qy	728	GGATTCACATCG-TGTG-----TTTCTATTTTTGAAGAAGCAGACATTTTAAAAAATGGT	781
Ddb	1088	CAACTCTCCGCGCAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC	1128
Qy	782	CACGTTTGGGTGCTCTTCAGATTCTTCAGG-----AAATTGCTTTGTATTGTA	828
Ddb	1129	TGCAGGTGAGGCTTCAGGGGATGCTGGGGCCCCACTGCCCTCGCTGCCCTGCCCTCCA	1188
Qy	829	T-ATTACAATGATC-----ACCG-----ACTGAG-----AATATTGTTT--TA	863
Ddb	1189	TCCTTCCTCTGTTTCTCTGTGGCGGGGACACACAGACACTGGGGGCTACCTCTTGGTTGATC	1248
Qy	864	CAATAGTTCCTGTGGGGCTGT-TTTTTTGTATTATAACCAATAAATTAGATGGTGAAAAA	921
Ddb	1249	CTCTGTGACTG-GGAGAGGTGCCCTTTGT--ATCCCCAATTA-----AAGGTAGAAAA	1298
RESULT 14			
ABO00408			
ID	ABO00408	standard; protein; 1300 AA.	
XX	AC	ABO00408;	
XX	AC	ABO00408;	
XX	XX	06-AUG-2003 (first entry)	
XX	XX	Novel human secreted and transmembrane PRO protein #2.	
XX	XX	Human; gene therapy; tumour necrosis factor alpha; TNF-alpha;	
XX	XX	chondrocyte stimulation; tumour; tissue typing.	
XX	OS	Homo sapiens.	
XX	OS	Homo sapiens.	
XX	XX	US2003032101-A1.	
XX	XX	PD	13-FEB-2003.
XX	XX	PP	17-JUN-2002; 2002US-00173695.
XX	XX	PP	18-SEP-1997; 97US-0059263P.
XX	XX	PP	18-SEP-1997; 97US-0059266P.
XX	XX	PP	17-OCT-1997; 97US-0062250P.
XX	XX	PP	21-OCT-1997; 97US-0063486P.
XX	XX	PP	24-OCT-1997; 97US-0063120P.
XX	XX	PP	24-OCT-1997; 97US-0063121P.
XX	XX	PP	28-OCT-1997; 97US-0063540P.
XX	XX	PP	28-OCT-1997; 97US-0063541P.
XX	XX	PP	28-OCT-1997; 97US-0063544P.
XX	XX	PP	29-OCT-1997; 97US-0063734P.
XX	XX	PP	31-OCT-1997; 97US-0063870P.
XX	XX	PP	31-OCT-1997; 97US-0064103P.
XX	XX	PP	13-NOV-1997; 97US-0065311P.
XX	XX	PP	21-NOV-1997; 97US-0066120P.
XX	XX	PP	24-NOV-1997; 97US-0066466P.
XX	XX	PP	24-NOV-1997; 97US-0066772P.
XX	XX	PP	11-DEC-1997; 97US-0069335P.
XX	XX	PP	12-DEC-1997; 97US-0069425P.
XX	XX	PP	17-DEC-1997; 97US-0069870P.
XX	XX	PP	18-DEC-1997; 97US-0068017P.
XX	XX	PP	10-MAR-1998; 98US-007450P.
XX	XX	PP	11-MAR-1998; 98US-0077632P.
XX	XX	PP	11-MAR-1998; 98US-0077649P.
XX	XX	PP	20-MAR-1998; 98US-0078886P.
XX	XX	PP	20-MAR-1998; 98US-0078939P.
XX	XX	PP	27-MAR-1998; 98US-0079664P.
XX	XX	PP	27-MAR-1998; 98US-0079786P.
XX	XX	PP	31-MAR-1998; 98US-0080107P.
XX	XX	PP	31-MAR-1998; 98US-0080194P.
XX	XX	PP	01-APR-1998; 98US-0080327P.
XX	XX	PP	01-APR-1998; 98US-0080333P.
XX	XX	PP	08-APR-1998; 98US-0081049P.
XX	XX	PP	08-APR-1998; 98US-0081070P.
XX	XX	PP	09-APR-1998; 98US-0081195P.

PR	01-JUL-1998;	98US-0091359P.	QY	1	GGGGCCCATTCCTGTTTCAGC-----CAGTCGCCAAG-----AATCATGAAGTCGCCAGTG	51
PR	01-JUL-1998;	98US-0091544P.	Db	40	GTGGCAGCGCGGTAGCAGCAATGGACTTTCCTCTGGGAACCCGCTTACGCTCTCCAGTG	99
PR	02-JUL-1998;	98US-0091478P.	QY	52	G--CAGCAC--CGCCACCGCCCGCGGC--CCAGACTGCCGCTGA--AGGCC-----	98
PR	02-JUL-1998;	98US-0091628P.	Db	100	GGACAGCGCATCGAAGAACCCACATGGCTCCCTGCAGAGCGAGACTGGGCCCTCAAC	159
PR	02-JUL-1998;	98US-0091632P.	QY	99	--GG-----CAAGACAG-----CGAGCG	114
PR	10-AUG-1998;	98US-0095982P.	Db	160	ATGGAGATCTCGACATCATCAACGAGCGGAGGAGTCCCAAGATGCCCTCCGAGCA	219
PR	17-AUG-1998;	98US-0096757P.	QY	115	GT-----GCGGG-----CGAGTGGTGC--GCTGT--CTGTC	142
PR	17-AUG-1998;	98US-0096867P.	Db	220	GTAAGAAGAGAAATCGTGGGGAATAAGAACTTCCACGAGGTGATGCTGCTCTCAGTTC	279
PR	17-AUG-1998;	98US-0096897P.	QY	143	TGAGCA-----GAGCGTGCCATCTCGCGCT-----GCGGGGC	176
PR	18-AUG-1998;	98US-0096949P.	Db	280	TTAGAAACCTGTGTCAAGAACTCGGGGCACCGCTTCCACGTCTGGTGGCCACGCGAGAC	339
PR	18-AUG-1998;	98US-0098552P.	QY	177	GCGGGGCGCG-----CCTGCTGCCCTGCTGGACGAGCAGCAGGTAAACG	222
PR	26-AUG-1998;	98US-0097954P.	Db	340	TTCTGTGAGAGTGTGCTGGTGGAGCAATCTCTGCC--AAGAACCAACCCACCACCATCG	397
PR	26-AUG-1998;	98US-0098014P.	QY	223	TGC-----TGCTCTACGACAT--GAAC--GGCT-----GTTACTCA--CGCC--	258
PR	01-SEP-1998;	98US-0098716P.	Db	398	TGATGACAAAGTGTCTCACTCATCTGCTGGGTGACCGCTTCCGACGTGCGCCG	457
PR	02-SEP-1998;	98US-0098803P.	QY	259	-TC---AAGAGCTGTGTGCCA-----CCCTGC-----CCC--AGAACCGC	293
PR	02-SEP-1998;	98US-0098821P.	Db	458	ATCTGACAGTG--TGCTCAACCATCTATGAGGACCTGCGGAGGAAGGCCCTGGAGTTC	516
PR	02-SEP-1998;	98US-0098843P.	QY	294	AAGG-----TGAGCAAGTGGAGATTCTCCA-----GCACGTATCTGACT--	333
PR	09-SEP-1998;	98US-0099602P.	Db	517	ATGACTGACCTGGACATGTGTCAACCATCCACACCCAGAGGACCGTGTCACTCAG	576
PR	10-SEP-1998;	98US-0099741P.	QY	334	---ACATCAGGAC---CTTCAGTTGGAGCTGAATCGGAATCGGAATGGGACCC	386
PR	10-SEP-1998;	98US-0099754P.	Db	577	AGACAAATCAGGACAGGATTCTGTGGGCATG-CTC--CAGCCAGCAAGAGACTCT-	632
PR	16-SEP-1998;	98US-0100662P.	QY	387	GCGGGC--GAGGCTCGGGTCCGGCTCGGCTCAGCAC--CCTCAACGG--CGAGATC	440
PR	16-SEP-1998;	98US-0100664P.	Db	633	---GGCAGCATGCTGCCCTCTG---CCGCCGCCGCCATCTCTCCGGTGCACGCC	686
PR	16-SEP-1998;	98US-0101751P.	QY	441	AGCGCCCTGACGGCCGAGGGCGCATGCTTCTCTCGC-----GACGATCGC-----	485
PR	17-SEP-1998;	98US-0100683P.	Db	687	ATAGCACCAAC--CCCGMAACAGATTGGGAAGCTGCGCAGTGAAGTGGATGGTGGTGG	745
PR	17-SEP-1998;	98US-0100684P.	QY	486	-ATCTTG-----TGTC-----GCTGA---AGC---GCCTCCCGCAG--GGACCGG	521
PR	23-SEP-1998;	98US-0101477P.	Db	746	GAACGTGAGGGTGTATGTGGAGATGTGTGACGGAGTGGTGCC--CACCCAGGCCGAGCCCG	804
PR	23-SEP-1998;	98US-0101479P.	QY	522	CGGACC-----CC-AGCCATCCAG-----GGGGCA	545
PR	24-SEP-1998;	98US-0101738P.	Db	805	CAGACCTGGAGCTGTGCGAGGACTCAACCGCACGTGCCGAGCCATGACAGCAGCGGTCC	864
PR	24-SEP-1998;	98US-0101739P.	QY	546	AGAGGAATTACGTCTGTGTGGTGC--TCCCCAACGCGCC--TC---GCCGATCTGAG	598
PR	29-SEP-1998;	98US-0102240P.	Db	865	TGAGTGATACCTGCTC--CGGGCCCATGCCCCAAGGAGCCCTTCAGAGCCCACTGCC	922
PR	29-SEP-1998;	98US-0102330P.	QY	599	GGAGAACAGACCGATC---GGC-GGCCACTG-----CG-----CCC	631
PR	30-SEP-1998;	98US-0102487P.	Db	923	AG---TCAGGCTGGGTGGAGGCTGGCCACAGTGGAAATTCGCCGAGCCTATTGTC	979
PR	01-OCT-1998;	98US-0102570P.	QY	632	TTAACTGCATC--CAGCTGGG--CTGAGGTGAGCACTGGC-----	671
PR	02-OCT-1998;	98US-0102687P.	Db	980	TACCTGCTGCTGTGATGGGGCCCATGGCTTTGG--CTGGCCACTGAGGTAGGGTG	1036
PR	06-OCT-1998;	98US-0103258P.	QY	672	--GAGGAGGGGCTCTCTCTGCACACCTACTAGTC--ACCAGAGACTTTTGGGGGTG	727
PR	06-OCT-1998;	98US-0103449P.	Db	1037	TGGAGGTGTGAGGCCCC---CTGAGGAGTGC--GGCGGCCAGGTTAC---GAGCTG	1087

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred. No. 3.5e-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

Qy	782	CACGTTTGGTGTCTTCAGATTCTGAGG-----AAATTGCTTTGTATTGTA	828
Db	1129	TCACGTCAGGCTTCAGGGGATGCTGGGGCCCACTGCCCTCCGCTGCCCTCCCA	1188
Qy	829	T-ATTACAATGATC-----ACCG-----ACTCAG-----AATATTGTTT--TA	863
Db	1189	TCCTTCCTCTGTTCTCTCTGGCGGGCCACCACGACCTGGGCTCACCTCTTGGTTGATC	1248
Qy	864	CAATAGTCTCTGGGGCTGT-TTTTTTGTATTAAACAATAATTTAGATGCTGAAAA	921
Db	1249	CTCTGTACTG-GGAGAGGTGCTTTTGT--ATCCCCATTA-----AAGGTAGAAAA	1298
	RESULT 13		
	ABU85269		
ID	ABU85269	standard; protein; 1300 AA.	
XX	AC		
XX	ABU85269;		
DT	30-JUN-2003	(first entry)	
DE	XX	Novel human secreted and transmembrane protein #2.	
KW	XX	Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy;	
KW	XX	chondrocyte stimulator; chromosome mapping; gene mapping;	
KW	XX	transgenic animal; knock-out animal; tumour.	
XX	XX		
OS	XX	Homo sapiens.	
PN	XX		
PD	US2003032114-A1.		
PF	13-FEB-2003.		
XX	20-JUN-2002; 2002US-00176919.		
XX	18-SEP-1997;	97US-0059263P.	
XX	18-SEP-1997;	97US-0059266P.	
XX	17-OCT-1997;	97US-0062250P.	
XX	21-OCT-1997;	97US-0063486P.	
XX	24-OCT-1997;	97US-0063120P.	
XX	24-OCT-1997;	97US-0063121P.	
XX	28-OCT-1997;	97US-0063340P.	
XX	28-OCT-1997;	97US-0063341P.	
XX	28-OCT-1997;	97US-0063344P.	
XX	28-OCT-1997;	97US-0063356P.	
XX	29-OCT-1997;	97US-0063734P.	
XX	31-OCT-1997;	97US-0063870P.	
XX	31-OCT-1997;	97US-0064103P.	
XX	31-NOV-1997;	97US-0065311P.	
XX	21-NOV-1997;	97US-0066120P.	
XX	24-NOV-1997;	97US-0066456P.	
XX	24-NOV-1997;	97US-0066772P.	
XX	11-DEC-1997;	97US-0069335P.	
XX	12-DEC-1997;	97US-0069425P.	
XX	17-DEC-1997;	97US-0069870P.	
XX	18-DEC-1997;	97US-0068017P.	
XX	10-MAR-1998;	98US-0077450P.	
XX	11-MAR-1998;	98US-0077632P.	
XX	11-MAR-1998;	98US-0077649P.	
XX	20-MAR-1998;	98US-0078986P.	
XX	20-MAR-1998;	98US-0078939P.	
XX	27-MAR-1998;	98US-0079656P.	
XX	31-MAR-1998;	98US-0079786P.	
XX	31-MAR-1998;	98US-0080107P.	
XX	31-MAR-1998;	98US-0080194P.	
XX	01-APR-1998;	98US-0080327P.	
XX	01-APR-1998;	98US-0080333P.	
XX	08-APR-1998;	98US-0081049P.	
XX	08-APR-1998;	98US-0081070P.	
XX	09-APR-1998;	98US-0081195P.	
XX	15-APR-1998;	98US-0081838P.	
XX	21-APR-1998;	98US-0082568P.	
XX	21-APR-1998;	98US-0082569P.	
XX	22-APR-1998;	98US-0082704P.	
XX	22-APR-1998;	98US-0082722P.	
XX	28-APR-1998;	98US-0083327P.	
XX	29-APR-1998;	98US-0083495P.	
XX	29-APR-1998;	98US-0083496P.	
XX	29-APR-1998;	98US-0083499P.	
XX	29-APR-1998;	98US-0083559P.	
XX	05-MAY-1998;	98US-0084366P.	
XX	05-MAY-1998;	98US-0084414P.	
XX	07-MAY-1998;	98US-0084463P.	
XX	07-MAY-1998;	98US-0084643P.	
XX	15-MAY-1998;	98US-0085579P.	
XX	15-MAY-1998;	98US-0085580P.	
XX	15-MAY-1998;	98US-0085582P.	
XX	15-MAY-1998;	98US-0085700P.	
XX	18-MAY-1998;	98US-0086023P.	
XX	22-MAY-1998;	98US-0086392P.	
XX	22-MAY-1998;	98US-0086486P.	
XX	28-MAY-1998;	98US-0087098P.	
XX	28-MAY-1998;	98US-0087208P.	
XX	02-JUN-1998;	98US-0087609P.	
XX	02-JUN-1998;	98US-0087753P.	
XX	03-JUN-1998;	98US-0087827P.	
XX	04-JUN-1998;	98US-0088025P.	
XX	04-JUN-1998;	98US-0088028P.	
XX	04-JUN-1998;	98US-0088033P.	
XX	04-JUN-1998;	98US-0088032P.	
XX	04-JUN-1998;	98US-0088167P.	
XX	05-JUN-1998;	98US-0088176P.	
XX	05-JUN-1998;	98US-0088212P.	
XX	05-JUN-1998;	98US-0088217P.	
XX	09-JUN-1998;	98US-0088655P.	
XX	10-JUN-1998;	98US-0088722P.	
XX	10-JUN-1998;	98US-0088738P.	
XX	10-JUN-1998;	98US-0088740P.	
XX	10-JUN-1998;	98US-0088811P.	
XX	10-JUN-1998;	98US-0088824P.	
XX	10-JUN-1998;	98US-0088825P.	
XX	10-JUN-1998;	98US-0088826P.	
XX	11-JUN-1998;	98US-0088861P.	
XX	11-JUN-1998;	98US-0088863P.	
XX	11-JUN-1998;	98US-0088876P.	
XX	12-JUN-1998;	98US-0089090P.	
XX	12-JUN-1998;	98US-0089105P.	
XX	16-JUN-1998;	98US-00895112P.	
XX	16-JUN-1998;	98US-0089514P.	
XX	17-JUN-1998;	98US-0089538P.	
XX	17-JUN-1998;	98US-0089598P.	
XX	17-JUN-1998;	98US-0089653P.	
XX	18-JUN-1998;	98US-0089908P.	
XX	19-JUN-1998;	98US-0089952P.	
XX	22-JUN-1998;	98US-0090246P.	
XX	22-JUN-1998;	98US-0090252P.	
XX	22-JUN-1998;	98US-0090254P.	
XX	24-JUN-1998;	98US-0090429P.	
XX	24-JUN-1998;	98US-0090435P.	
XX	24-JUN-1998;	98US-0090444P.	
XX	24-JUN-1998;	98US-0090461P.	
XX	24-JUN-1998;	98US-0090535P.	
XX	24-JUN-1998;	98US-0090540P.	
XX	25-JUN-1998;	98US-0090676P.	
XX	25-JUN-1998;	98US-0090678P.	
XX	25-JUN-1998;	98US-0090688P.	
XX	25-JUN-1998;	98US-0090690P.	
XX	25-JUN-1998;	98US-0090694P.	
XX	25-JUN-1998;	98US-0090695P.	
XX	25-JUN-1998;	98US-0090696P.	
XX	26-JUN-1998;	98US-0090862P.	
XX	26-JUN-1998;	98US-0090863P.	
XX	26-JUN-1998;	98US-0091010P.	

Qy	829	T-ATTCAATGATC-----ACCG-----ACTGAG-----AAATTGTTT--TA	863
Db	1189	TCCTTCCTCTGTCTCTTGGCCGGGACCACAGCACTGGGCTCACCTCTGTGTTGATC	1248
Qy	864	CAATAGTCTCTGGGGCTGT-TTTTTTTGTTATTAAACAAATAATTAGATGGTAAAAA	921
Db	1249	CTCTTGTACTG-GGAGAGGTGCTTTTGT--ATCCCCAATTA-----AAGGTAGAAA	1298

RESULT 12

RESULTS 12
ABU91965
ID ABU91965 standard; protein; 1300 AA.

AC ABU91965:

DT 11-AUG-2003 (first entry)

XX DE Novel human secreted and transmembrane PRO protein #2.

XX Human; gene therapy; chromosome identification; tissue typing.

XX
OS
Homo sapiens.

XX PN US2003027277-A1

XX
PD 06-FEB-2003

XX PF 21-JUN-2002: 2002US-00176985.

XX	PR	18-SEP-1997	97US-0053263P
PR	PR	18-SEP-1997	97US-0053266P
PR	PR	17-OCT-1997	97US-0062250P
PR	PR	21-OCT-1997	97US-0063486P
PR	PR	24-OCT-1997	97US-0063120P
PR	PR	24-OCT-1997	97US-0063121P
PR	PR	28-OCT-1997	97US-0063540P
PR	PR	28-OCT-1997	97US-0063541P
PR	PR	28-OCT-1997	97US-0063544P
PR	PR	28-OCT-1997	97US-0063564P
PR	PR	29-OCT-1997	97US-0063734P
PR	PR	31-OCT-1997	97US-0063870P
PR	PR	31-OCT-1997	97US-0064103P
PR	PR	31-OCT-1997	97US-0065311P
PR	PR	13-NOV-1997	97US-0066120P
PR	PR	24-NOV-1997	97US-0066466P
PR	PR	24-NOV-1997	97US-0066772P
PR	PR	11-DEC-1997	97US-0069333P
PR	PR	17-DEC-1997	97US-0069870P
PR	PR	18-DEC-1997	97US-0069870P
PR	PR	10-MAR-1998	97US-0068017P
PR	PR	11-MAR-1998	97US-0077450P
PR	PR	11-MAR-1998	98US-0077632P
PR	PR	11-MAR-1998	98US-0077643P
PR	PR	20-MAR-1998	98US-0078886P
PR	PR	27-MAR-1998	98US-0078933P
PR	PR	27-MAR-1998	98US-0079664P
PR	PR	31-MAR-1998	98US-0079786P
PR	PR	27-MAR-1998	98US-0080107P
PR	PR	31-MAR-1998	98US-0080194P
PR	PR	01-APR-1998	98US-0080327P
PR	PR	01-APR-1998	98US-0080333P
PR	PR	08-APR-1998	98US-0081049P
PR	PR	08-APR-1998	98US-0081070P
PR	PR	09-APR-1998	98US-0081192P
PR	PR	15-APR-1998	98US-0081832P
PR	PR	21-APR-1998	98US-0082568P
PR	PR	21-APR-1998	98US-0082569P
PR	PR	22-APR-1998	98US-0082704P
PR	PR	22-APR-1998	98US-0082797P
PR	PR	28-APR-1998	98US-0083322P
PR	PR	28-APR-1998	98US-0083495P
PR	PR	29-APR-1998	98US-0083466P
PR	PR	29-APR-1998	98US-0083493P

PR	29-APR-1998	98US-008355599
PR	05-MAY-1998	98US-00843666P
PR	06-MAY-1998	98US-00844114P
PR	07-MAY-1998	98US-00846340P
PR	07-MAY-1998	98US-00846460P
PR	07-MAY-1998	98US-00846463P
PR	07-MAY-1998	98US-00846433P
PR	15-MAY-1998	98US-0085573P
PR	15-MAY-1998	98US-0085580P
PR	15-MAY-1998	98US-0085592P
PR	15-MAY-1998	98US-0085700P
PR	18-MAY-1998	98US-0086023P
PR	22-MAY-1998	98US-0086392P
PR	22-MAY-1998	98US-0086486P
PR	28-MAY-1998	98US-0087098P
PR	28-MAY-1998	98US-0087208P
PR	02-JUN-1998	98US-0087609P
PR	02-JUN-1998	98US-0087759P
PR	02-JUN-1998	98US-0087827P
PR	04-JUN-1998	98US-0088025P
PR	04-JUN-1998	98US-0088028P
PR	04-JUN-1998	98US-0088029P
PR	04-JUN-1998	98US-0088033P
PR	04-JUN-1998	98US-0088326P
PR	05-JUN-1998	98US-0088167P
PR	05-JUN-1998	98US-0088202P
PR	05-JUN-1998	98US-0088212P
PR	05-JUN-1998	98US-0088217P
PR	08-JUN-1998	98US-0088565P
PR	10-JUN-1998	98US-0088722P
PR	10-JUN-1998	98US-0088738P
PR	10-JUN-1998	98US-0088740P
PR	10-JUN-1998	98US-0088811P
PR	10-JUN-1998	98US-0088824P
PR	10-JUN-1998	98US-0088825P
PR	10-JUN-1998	98US-0088826P
PR	10-JUN-1998	98US-0088861P
PR	11-JUN-1998	98US-0088863P
PR	11-JUN-1998	98US-0088876P
PR	12-JUN-1998	98US-0089090P
PR	12-JUN-1998	98US-0089105P
PR	15-JUN-1998	98US-0089512P
PR	15-JUN-1998	98US-0089514P
PR	17-JUN-1998	98US-0089536P
PR	17-JUN-1998	98US-0089598P
PR	17-JUN-1998	98US-0089653P
PR	18-JUN-1998	98US-0089908P
PR	19-JUN-1998	98US-0089952P
PR	22-JUN-1998	98US-0090246P
PR	22-JUN-1998	98US-0090252P
PR	22-JUN-1998	98US-0090254P
PR	24-JUN-1998	98US-0090429P
PR	24-JUN-1998	98US-0090435P
PR	24-JUN-1998	98US-0090444P
PR	24-JUN-1998	98US-0090461P
PR	24-JUN-1998	98US-0090535P
PR	24-JUN-1998	98US-0090540P
PR	24-JUN-1998	98US-0090676P
PR	25-JUN-1998	98US-0090678P
PR	25-JUN-1998	98US-0090688P
PR	25-JUN-1998	98US-0090690P
PR	25-JUN-1998	98US-0090694P
PR	25-JUN-1998	98US-0090695P
PR	25-JUN-1998	98US-0090696P
PR	26-JUN-1998	98US-0090696P
PR	26-JUN-1998	98US-009105413
PR	26-JUN-1998	98US-0090862P
PR	26-JUN-1998	98US-0090863P
PR	01-JUL-1998	98US-0091010P
PR	01-JUL-1998	98US-0091359P
PR	01-JUL-1998	98US-0091544P
PR	02-JUL-1998	98US-0091478P
PR	02-JUL-1998	98US-0091486P
PR	02-JUL-1998	98US-0091626P
PR	02-JUL-1998	98US-0091628P

PR	24-JUL-1998;	98US-0094006P.	QY	99	--GG--	-----CAAGCAG-----	-----CGAGCG	114
PR	04-AUG-1998;	98US-0095282P.	Db	160	ATGAGATCTGCGACATCATCAACAGAGCGAGGAAGGTCCCAAGATGCCCTCCGAGCA	219		
PR	10-AUG-1998;	98US-0096012P.	QY	115	GT-----CGGG-----	-----CGAGGTGGTGC-----GCTGT-----CTGTC	142	
PR	17-AUG-1998;	98US-0096766P.	Db	220	GTAAGAGAGAAATCGTGGGAATAGAACTTCCACGAGGTGATGCTGGCTCTCA	279		
PR	17-AUG-1998;	98US-0096867P.	QY	143	TGAGCA-----	-----GAGCGTGCCCATCTGCGCT-----	-----GCCGGGCG	176
PR	17-AUG-1998;	98US-0096891P.	Db	280	TTAGAAACCTGTGTCAAGAACTCGGGGCAACCGCTCCACGCTGCTGGTCCGACGAGCAGAC	339		
PR	18-AUG-1998;	98US-0096959P.	QY	177	GCGGGGCGG-----	-----CCTGCTGCTGCTGCTGAGCAGCAGCAGGTAACG	222	
PR	26-AUG-1998;	98US-0097022P.	Db	340	TTCTGGAGTGTGCTGTGAGGACCATCTGCCC--	-----AAGAAACCAACCAACCCACCTCG	397	
PR	26-AUG-1998;	98US-0097355P.	QY	223	TGC-----TGCTCTACGACAT-----GAAC-----GGCT-----GTTATCTCA-----CGCC-----	258		
PR	26-AUG-1998;	98US-0097974P.	Db	398	TGCATGACAAAGTGTCTCAACCTCATCCAGTCTGGGCTGACGCGTTCGCGAGCTCGCCG	457		
PR	01-SEP-1998;	98US-0098014P.	QY	259	-TC-----AAGGAGCTGTGCCCCA-----	-----CCCTGC-----	-----CCC-----AGAACCGC	293
PR	01-SEP-1998;	98US-009812P.	Db	458	ATCTGACAGTG--TGCTACCATCTATGAGGACCTGCGGAGAAAGGCTGGAGTTCCTCC	516		
PR	01-SEP-1998;	98US-0098716P.	QY	294	ANGG-----TGAGCAAGTGGAGATTTCTCA-----	-----GACGTCATCTCACT---	333	
PR	02-SEP-1998;	98US-0098803P.	Db	517	ATGACTGACCTGGACATGCTGTCAACCATCCACACACACAGAGGACCGTGTTCAACTCAG	576		
PR	02-SEP-1998;	98US-0098821P.	QY	334	---ACATCAGGAC---CTTCAGTTGGAGCTGAATCGGAATCCGAAATTCGGACCCCG	386		
PR	02-SEP-1998;	98US-0100388P.	Db	577	AGACACATCAGGACAGGATTTCTGGGCACTG-ATCTC--	-----CAGCCAGCAGAGGACTCT-	632	
PR	02-SEP-1998;	98US-0100642P.	QY	387	GGGGGCC--GAGGGCTCCGGTCCGGCTCCGTCTCAGCAC--	-----CCTCAACGG-----CGAGATC	440	
PR	02-SEP-1998;	98US-0100684P.	Db	633	---GGGACAGCATGCTGCCCTCTG--	-----CCCGCCCGCCCATCTCTCGGTCACACGCCC	686	
PR	02-SEP-1998;	98US-0100930P.	QY	441	AGCGCCTGACGGCGGAGCGGCGATGCTGCTGCG--	-----GAGCATCGC-----	485	
PR	02-SEP-1998;	98US-0100849P.	Db	687	ATAGACCAAC--CCCGAACAGATTGGGAAGTGGCAGTGGAGCTGGAGATGCTGAGTGG	745		
PR	02-SEP-1998;	98US-0101042P.	QY	486	-ATCTTG-----TGTC-----GCTGA-----AGC-----GCCTCCCCCAG-----GAGCCGG	521		
PR	02-SEP-1998;	98US-0101471P.	Db	746	GAAAGTGGAGGTGATGTCGAGATGCTGACGAGTGTGTGTC--	-----CAGCCAGGCGGAGCCCG	804	
PR	02-SEP-1998;	98US-0101475P.	QY	522	CGGACC-----	-----CC-AGCCATCCAG-----GGGCA	545	
PR	02-SEP-1998;	98US-0101738P.	Db	805	CAGACCTGGAGCTGTGAGGAGCTCAACCGCAGCTGCGAGCCATGAGAGCGGCTCC	864		
PR	02-SEP-1998;	98US-0101739P.	QY	546	AGAGGAATACGTGCTGTGGTCTC--	-----TCCCCAACGGGCC--TC-----GCCGATCTGAG	598	
PR	02-SEP-1998;	98US-0102240P.	Db	865	TGAGTGATACCTGCTC--	-----CGGGCCCATGCCCCAAGGAGCCCTTCAGAGCCACACTGCC	922	
PR	02-SEP-1998;	98US-0102331P.	QY	599	GGAGAACAGACCGATC--	-----GGC--GGCCACTG--	-----CG-----	631
PR	02-SEP-1998;	98US-0102487P.	Db	923	AG--	-----TGAGGCTGCTGGAGGCTGGCCACAGTGGAAATTCTGCCGAGCCTATTGTC	979	
PR	02-SEP-1998;	98US-0102570P.	QY	632	TTAATCTGATC--	-----CAGCCTGGGG--	-----CTGAGGCTGAGGCACTGGC-----	671
PR	01-OCT-1998;	98US-0102684P.	Db	980	TACCTGCTCTGCTGATGGGCCCCCATGGCTTTGG--	-----CTGGCCACTGAGGGTAGGGTG	1036	
PR	06-OCT-1998;	98US-0103258P.	QY	672	--GAGGAGAGGGCGCTCTCTGACACCTACTAGTC--	-----ACCAGAGACTTTAGGGGGTG	727	
PR	06-OCT-1998;	98US-0103449P.	Db	1037	TGGAGGTGGAGGCCCC--	-----CTGAGGAGCTGC--	-----GGCGGCCCGAGTAC-----GAAGCTG	1087
PR	06-OCT-1998;	98US-0103449P.	QY	728	GGATTCACTCG--TGTC-----	-----TTTCTATTTTTTGAAGAGAGACATTTTAAAAATGTT	781	
PR	06-OCT-1998;	98US-0103258P.	Db	1088	CAACTCTGCGCGAGTGGGCGAGATCTCAT--	-----CAGCC-----	-----CCAGGC	1128
PR	06-OCT-1998;	98US-0103258P.	QY	782	CAGCTTTGGTGTCTTCAGATTCTTCTAGG--	-----	-----AAATTGCTTTTGTATTGTA	828
PR	06-OCT-1998;	98US-0103449P.	Db	1129	TGCAGGTGAGGCTTCAGGGGATGCTGGGCCCCCACTGCCCTCGCTGCTGCTCCCTCAAC	1188		

Query Match 43.2%; Score 2491; DB 6; Length 1300;
Best Local Similarity 42.6%; Pred. No. 3.5e-104;
Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;
QY 1 GGGGCCCCATCTGTTTTCAGC-----CAGTGGCCAAAG---AATCATGAAAGTGGCCAGTG 51
Db 40 GTGGCAGCGCGGTAGCAGCAATGGAATTTCTCTCTGGGGAACCCGTTTCACTCTCCAGTG 99
QY 52 G--CAGCAC--CGCCACCGCGCGGCG--CCGAGCTGCGGCTGA--AGGCC----- 98
Db 100 GGACAGCGCATCGAAGAACCAAGATGCTCTCTGAGAGCAGGAGCTGGGCCCCCTCAAC 159

QY 864 CAATAGTCTCTGGGCGCTG-TTTTTCGTATTAAACAATAAATTAGATGGTGA AAAA 921
Db 1249 CTCCTGTACTG-GGAGAGGTGCTTTTGT--ATCCCAATTA-----AAGGTAGAAAA 1298

RESULT 11
ABU98259
ID ABU98259 standard; protein; 1300 AA.
XX AC ABU98259;
XX DT 30-JUL-2003 (first entry)
XX DE Novel human secreted and transmembrane protein #2.
XX KW Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy;
KW chondrocyte stimulator; tumour; adrenal tumour; lung tumour;
KW colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; chromosome identification.
XX OS Homo sapiens.
XX PN US2003017544-A1.
XX PD 23-JAN-2003.
XX PF 21-JUN-2002; 2002US-00176915.
XX PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069135P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.

05-MAY-1998; 98US-0084366P.
06-MAY-1998; 98US-0084414P.
07-MAY-1998; 98US-0084639P.
07-MAY-1998; 98US-0084640P.
07-MAY-1998; 98US-0084643P.
15-MAY-1998; 98US-0085579P.
15-MAY-1998; 98US-0085580P.
15-MAY-1998; 98US-0085582P.
15-MAY-1998; 98US-0085700P.
18-MAY-1998; 98US-0086023P.
22-MAY-1998; 98US-0086392P.
22-MAY-1998; 98US-0086486P.
28-MAY-1998; 98US-0087098P.
28-MAY-1998; 98US-0087208P.
02-JUN-1998; 98US-0087609P.
03-JUN-1998; 98US-0087759P.
04-JUN-1998; 98US-0087827P.
04-JUN-1998; 98US-0088025P.
04-JUN-1998; 98US-0088028P.
04-JUN-1998; 98US-0088029P.
04-JUN-1998; 98US-0088033P.
04-JUN-1998; 98US-0088126P.
05-JUN-1998; 98US-0088167P.
05-JUN-1998; 98US-0088202P.
05-JUN-1998; 98US-0088212P.
05-JUN-1998; 98US-0088217P.
09-JUN-1998; 98US-0088555P.
10-JUN-1998; 98US-0088722P.
10-JUN-1998; 98US-0088738P.
10-JUN-1998; 98US-0088740P.
10-JUN-1998; 98US-0088811P.
10-JUN-1998; 98US-0088824P.
10-JUN-1998; 98US-0088825P.
10-JUN-1998; 98US-0088826P.
11-JUN-1998; 98US-0088861P.
11-JUN-1998; 98US-0088863P.
11-JUN-1998; 98US-0088876P.
12-JUN-1998; 98US-0089090P.
12-JUN-1998; 98US-0089105P.
16-JUN-1998; 98US-0089512P.
16-JUN-1998; 98US-0089514P.
17-JUN-1998; 98US-0089538P.
17-JUN-1998; 98US-0089598P.
17-JUN-1998; 98US-0089653P.
18-JUN-1998; 98US-0089908P.
19-JUN-1998; 98US-0089952P.
22-JUN-1998; 98US-0090246P.
22-JUN-1998; 98US-0090252P.
22-JUN-1998; 98US-0090254P.
24-JUN-1998; 98US-0090429P.
24-JUN-1998; 98US-0090435P.
24-JUN-1998; 98US-0090444P.
24-JUN-1998; 98US-0090461P.
24-JUN-1998; 98US-0090535P.
24-JUN-1998; 98US-0090540P.
25-JUN-1998; 98US-0090676P.
25-JUN-1998; 98US-0090678P.
25-JUN-1998; 98US-0090688P.
25-JUN-1998; 98US-0090690P.
25-JUN-1998; 98US-0090694P.
25-JUN-1998; 98US-0090695P.
25-JUN-1998; 98US-0090696P.
26-JUN-1998; 98US-00105413.
26-JUN-1998; 98US-0090862P.
26-JUN-1998; 98US-0090863P.
26-JUN-1998; 98US-0091010P.
01-JUL-1998; 98US-0091359P.
01-JUL-1998; 98US-0091544P.
02-JUL-1998; 98US-0091478P.
02-JUL-1998; 98US-0091486P.
02-JUL-1998; 98US-0091626P.
02-JUL-1998; 98US-0091628P.
02-JUL-1998; 98US-0091632P.

PR	04-AUG-1998;	98US-0095282P.	
PR	10-AUG-1998;	98US-0095998P.	
PR	10-AUG-1998;	98US-0096012P.	
PR	17-AUG-1998;	98US-0096757P.	
PR	17-AUG-1998;	98US-0096766P.	
PR	17-AUG-1998;	98US-0096867P.	
PR	17-AUG-1998;	98US-0096891P.	
PR	17-AUG-1998;	98US-0096897P.	
PR	18-AUG-1998;	98US-0096949P.	
PR	18-AUG-1998;	98US-0096959P.	
PR	18-AUG-1998;	98US-0097022P.	
PR	18-AUG-1998;	98US-0097952P.	
PR	26-AUG-1998;	98US-0097954P.	
PR	26-AUG-1998;	98US-0097955P.	
PR	26-AUG-1998;	98US-0097971P.	
PR	26-AUG-1998;	98US-0097974P.	
PR	26-AUG-1998;	98US-0098014P.	
PR	01-SEP-1998;	98US-0098716P.	
PR	01-SEP-1998;	98US-0098723P.	
PR	02-SEP-1998;	98US-0098803P.	
PR	02-SEP-1998;	98US-0098821P.	
PR	02-SEP-1998;	98US-0098843P.	
PR	09-SEP-1998;	98US-0099602P.	
PR	10-SEP-1998;	98US-0099741P.	
PR	10-SEP-1998;	98US-0099754P.	
PR	10-SEP-1998;	98US-0099763P.	
PR	10-SEP-1998;	98US-0099812P.	
PR	15-SEP-1998;	98US-0100388P.	
PR	16-SEP-1998;	98US-0100662P.	
PR	16-SEP-1998;	98US-0100664P.	
PR	16-SEP-1998;	98US-0101751P.	
PR	16-SEP-1998;	98WO-US019330.	
PR	17-SEP-1998;	98US-0100683P.	
PR	17-SEP-1998;	98US-0100864P.	
PR	17-SEP-1998;	98US-0100919P.	
PR	17-SEP-1998;	98US-0100930P.	
PR	18-SEP-1998;	98US-0100949P.	
PR	18-SEP-1998;	98US-0101014P.	
PR	18-SEP-1998;	98US-0101068P.	
PR	23-SEP-1998;	98US-0101471P.	
PR	23-SEP-1998;	98US-0101472P.	
PR	23-SEP-1998;	98US-0101475P.	
PR	23-SEP-1998;	98US-0101477P.	
PR	24-SEP-1998;	98US-0101738P.	
PR	24-SEP-1998;	98US-0101739P.	
PR	24-SEP-1998;	98US-0101743P.	
PR	24-SEP-1998;	98US-0101922P.	
PR	25-SEP-1998;	98US-0101766P.	
PR	29-SEP-1998;	98US-0102207P.	
PR	29-SEP-1998;	98US-0102240P.	
PR	29-SEP-1998;	98US-0102330P.	
PR	29-SEP-1998;	98US-0102331P.	
PR	30-SEP-1998;	98US-0102487P.	
PR	30-SEP-1998;	98US-0102570P.	
PR	30-SEP-1998;	98US-0102571P.	
PR	01-OCT-1998;	98US-0102684P.	
PR	01-OCT-1998;	98US-0102687P.	

Query Match 43.2%; Score 2491; DB 6; Length 1300;

Best Local Similarity 42.6%; Pred. No. 3.5e-104;

Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

Qy	1	GGGGCCCATCTGTTTCAGC-----CAGTGGCAAG---AATCATGAAAGTCGCCAGTG 51
Db	40	GTGGCAGCGCGGTAGCAGCAATGGACTTTCTCTGGGGAACCCGTTTCAGCTCTCCAGTG 99
Qy	52	G--CAGCAC--CGCCACCGCCCGCGGGC--CCAGCTGCGCGCTGA--AGGCC----- 98
Db	100	GGACAGCGCATCGAAGAACCCACAGATGGCTCCCTGACAGCGAGGACTGGGCCCTCAAC 159
Qy	99	--GG-----CAAGACAG-----CGAGCG 114
Db	160	ATGGAGATCTGCGATCATCAACGAGCGGAGGAAGTCCCAAGATGCCCTCCGAGCA 219

Qy	115	GT-----GCGGG-----CGAGGTGGTGC--GCTGT--CTGTC 142
Db	220	GTAAAGAGAGAAATCGTCGGGAATAAGAACTTCCACAGGTGATGCTGCTCTCACAGTC 279
Qy	143	TGAGCA-----GAGCGTGGCATCTCGCGT-----GCCGGGCG 176
Db	280	TTAGAAACCTGTGTCAAGAACTCGGGGCACCGTCCACGTGTGTGTGTCACAGCAGGAC 339
Qy	177	GCCGGGGCGCG-----CCTGCTGCTGCTGTGACGAGCAGCAGGAGTAAACG 222
Db	340	TTCTGTGAGAGTGTGCTGTGAGGACCATCTGCCCC--AAGAACAAACCCACCCACCATCG 397
Qy	223	TGC-----TGCTTACGACAT--GAAC--GGCT-----GTTACTCA--CGCC-- 258
Db	398	TGCATGACAAAGTGTCAACCTCATCCAGTCTGGGCTGACGGGTTCCGAGCTCGCCCG 457
Qy	259	-TC---AAGGAGCTGTGCCCCA-----CCCTGC-----CCC--AGAACCGC 293
Db	458	ATCTGACAGGTG-TGGTCAACCATCTATGAGGACCTCGGAGGAAGAGCCCTGGAGTTCCCC 516
Qy	294	AAG-----TGAGCAAGTGGAGATCTCCA-----GCACGTCAATCGACT-- 333
Db	517	ATGACTGACCTGACATGTGTCAACCATCCACACCCAGAGGACCGTGTTCATCTCAG 576
Qy	334	---ACATCAGGAC---CTTCAGTTGGAGCTGAATCGGAATCCGAAAGTTGGGACCCCC 386
Db	577	AGACACAATCAGGACAGGATTTCTGTGGCACTG-ACTC--CAGCCAGCAAGAGACTCT- 632
Qy	387	GGGGGCC-GAGGGCTGCGGTCCGGCTCGGCTCAGCAC--CCTCAACGG---CGAGATC 440
Db	633	---GGCCAGCATGCTGCCCTCTG---CCGCCCCCCTCATCTCTCCGCTGACAGCCC 686
Qy	441	AGCGCCTGACGGCCGAGGCGGCATGCTTCTCTGCG-----GACGATCGC----- 485
Db	687	ATAGCACCAAC-CCCGAAGCAGATTGGGAAGCTGCGCAGTGAGTGGAGATGCTGAGTGG 745
Qy	486	-ATCTTG-----TGTC-----GCTGA--AGC-----GCTTCCCCCAG--GGACCGG 521
Db	746	GAACGTGAGGCTGATGTCGGAGATGCTGACGAGCTGCTGCGC--CACCAGCGCCAGCCCG 804
Qy	522	CGGACC-----CC-AGCCATCCAG-----GGGGCA 545
Db	805	CAGACCTGGAGCTGCTCAGGAGCTCAACCGCAGCTGCCGAGCCATCAGCAGCGGCTCC 864
Qy	546	AGAGAAATTAGTCTCTGTGGGTC--TCCCCAACCGCGC--TC---GCCGATCTGAG 598
Db	865	TGAGTGATACCTCTGCTC--CGGGCCCATGCCCCAAGGAGCCCTTACAGAGCCACACTGCC 922
Qy	599	GGAGAACAAAGACCGATC---GGC-GGCCACTG-----CG-----CCC 631
Db	923	AG---TCGAGCCCTGGCTGGAGGCTGCCACAGTGGAAATTCGCCGAGCCTATTGTCCC 979
Qy	632	TTAACTGCATC--CAGCCTGGGG--CTGAGGCTGAGGCACTGGC----- 671
Db	980	TACCTGC-TCTGCTGATGGGGCCCATGGCTTTGG--CTGGCCACTGAGGGTAGGGTG 1036
Qy	672	--GAGGAGGGCGCTCTCTCTCACACCTACTAGTC--ACCAGAGACTTTAGGGGCTG 727
Db	1037	TGGAGGTGTGAGGCCCCC---CTGAGGAGCTGC--GGCGGCCACAGGTAC---GAAGCTG 1087
Qy	728	GGATTCCACTCG-TGTG-----TTTCTATTTTTGAAGAGCAGACATTTAAAAATGGT 781
Db	1088	CAACTCTCGCGCAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC 1128
Qy	782	CACGTTTGGTGTCTCAGATTTCTGAGG-----AAATTGCTTTGTATTGTA 828
Db	1129	TGCAGGTGAGCTTCAGGGGATGCTGGGGCCCACTGCCCTCGCTGCTTGGCCCTCCA 1188
Qy	829	T-ATTACAATGATC-----ACCG-----ACTGAG-----AATATGTTT--TA 863
Db	1189	TCCITCTCTGTCTCTTCTGSCCGGGGACACAGCACTGGGGCTCACCTCTTGTGTTGATC 1248

RESULT 10
ABU99044
ID ABU99044 standard; protein; 1300 AA.
XX
XX AC ABU99044;
XX
XX DT 01-AUG-2003 (first entry)
XX
XX DE Novel human secreted and transmembrane protein #2.
XX
XX KW Human; secreted and transmembrane protein; PRO; cytostatic; gene therapy;
KW chondrocyte stimulator; tumour; adrenal tumour; lung tumour;
KW colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; TNF-alpha release;
KW tumour necrosis factor alpha release; chondrocyte cell proliferation;
KW chondrocyte cell differentiation; pharmaceutical; diagnostic; biosensor;
KW bio-reactor.
XX
XX OS Homo sapiens.
XX
XX PN US2003013153-A1.
XX
XX PD 16-JAN-2003.
XX
XX PF 19-JUN-2002; 2002US-00175737.
XX
XX PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0069017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078866P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 09-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 05-MAY-1998; 98US-0084366P.
PR
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084639P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 15-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087098P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 09-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091646P.
PR 02-JUL-1998; 98US-0091648P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.

PR	17-AUG-1998;	98US-0096897P.	Query Match	43.2%;	Score 2491;	DB 6;	Length 1300;
PR	18-AUG-1998;	98US-0096949P.	Best Local Similarity	42.6%;	Pred. No. 3.5e-104;		
PR	18-AUG-1998;	98US-0096959P.	Matches 562;	Conservative	0;	Mismatches 299;	Indels 458; Gaps 83;
PR	18-AUG-1998;	98US-0097022P.					
PR	26-AUG-1998;	98US-0097952P.					
PR	26-AUG-1998;	98US-0097954P.					
PR	26-AUG-1998;	98US-0097955P.					
PR	26-AUG-1998;	98US-0097971P.					
PR	26-AUG-1998;	98US-0097974P.					
PR	01-SEP-1998;	98US-0098014P.					
PR	01-SEP-1998;	98US-0098716P.					
PR	01-SEP-1998;	98US-0098723P.					
PR	02-SEP-1998;	98US-0098803P.					
PR	02-SEP-1998;	98US-0098821P.					
PR	02-SEP-1998;	98US-0098843P.					
PR	02-SEP-1998;	98US-0099602P.					
PR	10-SEP-1998;	98US-0099741P.					
PR	10-SEP-1998;	98US-0099754P.					
PR	10-SEP-1998;	98US-0099812P.					
PR	15-SEP-1998;	98US-0100388P.					
PR	16-SEP-1998;	98US-0100622P.					
PR	16-SEP-1998;	98US-0100664P.					
PR	16-SEP-1998;	98US-0101751P.					
PR	16-SEP-1998;	98US-0101751P.					
PR	17-SEP-1998;	98US-0100683P.					
PR	17-SEP-1998;	98US-0100684P.					
PR	17-SEP-1998;	98US-0100919P.					
PR	17-SEP-1998;	98US-0100930P.					
PR	18-SEP-1998;	98US-0100849P.					
PR	18-SEP-1998;	98US-0101044P.					
PR	18-SEP-1998;	98US-0101068P.					
PR	23-SEP-1998;	98US-0101471P.					
PR	23-SEP-1998;	98US-0101472P.					
PR	23-SEP-1998;	98US-0101475P.					
PR	23-SEP-1998;	98US-0101477P.					
PR	24-SEP-1998;	98US-0101738P.					
PR	24-SEP-1998;	98US-0101739P.					
PR	24-SEP-1998;	98US-0101743P.					
PR	24-SEP-1998;	98US-0101922P.					
PR	25-SEP-1998;	98US-0101786P.					
PR	29-SEP-1998;	98US-0102207P.					
PR	29-SEP-1998;	98US-0102240P.					
PR	29-SEP-1998;	98US-0102330P.					
PR	29-SEP-1998;	98US-0102331P.					
PR	30-SEP-1998;	98US-0102487P.					
PR	30-SEP-1998;	98US-0102570P.					
PR	30-SEP-1998;	98US-0102571P.					
PR	01-OCT-1998;	98US-0102684P.					
PR	01-OCT-1998;	98US-0102687P.					
PR	02-OCT-1998;	98US-0102965P.					
PR	06-OCT-1998;	98US-0103258P.					
PR	06-OCT-1998;	98US-0103449P.					
QY	1	GGGGCCATTCTGTTTCAGC-----CAGTGGCCAAAG---AA7CATGAAGTGGCCAGTG 51					
Db	40	GTGGCAGCGGGGTAGCAGCAATGGAATTTCTCTGGGGAACCCGTTCACTCTCCAGTG 99					
QY	52	G---CAGCAC---CGCACCGCCCGCGGCG---CCAGCTGGCGCTGA---AGACC----- 98					
Db	100	GGACAGCGGATCGAAGAACGACAGATGGCTCTCTGAGAGCGAGGACTGGGCGCTCAAC 159					
QY	99	--GG-----CAAGACAG----- 114					
Db	160	ATGGAGATCTGGACATCATCAAGACGAGGAGGAGTCCCAAGATGCCCTCCAGCA 219					
QY	115	GT-----GCGGG-----CGAGTGGTGC---GCTG---CTGTC 142					
Db	220	GTAAGAGAGAGATCTGGGGAATAGAACTTCCACAGGATGATGCTGCTCTCACAGTC 279					

QY	143	TGAGCA-----GAGCGTGGCCATCTCGGCT-----GCGGGGC 176
Db	280	TTAGAAACCTGTGTCAAGAACTGCGGGCACCGCTTCCACGCTGCTGGTGGCCAGGAC 339
QY	177	GCCGGGGCGG-----CTGCTGCGCTCTGACGAGCAGCAGGTAACG 222
Db	340	TTCTGTGGAGAGTGTGCTGTGAGACCATCTCTGCCCC--AAGAACAAACCCACCACTCG 397
QY	223	TGC-----TGCTCTACGACAT---GAAC---GGCT---GTTACTCA---GCGC--- 258
Db	398	TGCATGACAAAGTGTCTCAACCTCATCCAGTCTGGGCTGAGCGTTCGCGAGTCTGCGCG 457
QY	259	-TC---AAGGAGCTGGTGGCA-----CCCTGC-----CCC--AGAACGCG 293
Db	458	ATCTGACAGGTG--TGCTACCATCTATAGAGACCTGCGGAGAAAGGCTCGAGTTCGCC 516
QY	294	AAGG-----TGAGCAAGGTGAGATTTCTCCA-----GACGTCTACTGACT--- 333
Db	517	ATGACTGACCTGGACATGTGTTCACCCATCCACACCCAGAGGACCGTGTCAACTCAG 576
QY	334	---ACATCAGGAC---CTTCAGTGGAGCTGACTCGGATCCGAACTTGGGACCCCC 386
Db	577	AGACACAATCAGGACAGGATTTCTGGGCACTG-ACTC--CAGCCAGCAAGAGGACTCT- 632
QY	387	GGGGCC--GAGGGCTCGGCTCCGGCTCCGCTCTAGCAC--CCTCAACGG---CGAGATC 440
Db	633	---GGCCAGCATGCTGCCCTCTG---CCGCGCCCGCCATACTCTCCGGTGACACGCC 686
QY	441	AGCGCCTGAGCGCGAGGGCGCATGCTTCCTGG---GACGATCGC----- 485
Db	687	ATAGCACCAAC--CCGGAACAGATTGGGAGCTGCGAGTGGAGTGGAGATGGTAGTGG 745
QY	486	-ATCTTG-----TGTC-----GCTGA---AGC---GCTCCCCCAG---GGACGG 521
Db	746	GAACGTGAGGGTGATGTGCGAGATGTGACGGAGTGTGTGCC--CACCCAGGCGAGCGCG 804
QY	522	CGGACC-----CC-AGCCATCCAG---GGGCA 545
Db	805	CAGACTGGAGCTGCTGAGGAGCTCAACCGCACCTGCGAGCCATGACAGCGGGTCC 864
QY	546	AGAGGAATACGTGCTCTGTGGTC--TCCCCACAGCGGC--TC---GCGGATCTGAG 598
Db	865	TGAGTGATACCTGTCTC--CGGGCCCATGCCCCAAGAGAGCCCTTCAGAGCCACACTGCC 922
QY	599	GGAGAACAGACCGATC---GGC--GGCCACTG---CG-----CCC 631
Db	923	AG---TCGAGGCTGGCTGGAGGCTGGCCACAGTGGAAATCTGCGGAGCCTATTGTCCC 979
QY	632	TTAACTGCATC--CAGCCTGGGG--CTGAGGCTGAGGACATGCGC----- 671
Db	980	TACCCCTGC-TCTGCTGATGGGGCCCATGGGCTTTGG--CTGGCCACTGAGGGTAGGGTG 1036
QY	672	--GAGGAGGGGGCTCTCTGACACCTACTACTAGTC--ACCAGAGACTTTAGGGGGTG 727
Db	1037	TGAGGTGTGAGGCCCC--CTGAGGAGCTGC--GGGGGCCAGGTAC---GAAGCTG 1087
QY	728	GGATTCACCTCG-TGTG-----TTTCTATTTTGAAGAGCAGACATTTTAAAAAATGTT 781
Db	1088	CAACTCTGCGCGCAGTGGCGGAGATCTCAT-----CAGCC-----CCAGGC 1128
QY	782	CAGTTTGGTGTCTCAGATTCTCAGG-----AAATTGCTTTGTTATGTA 828
Db	1129	TGAGGTGAGCTTTCAGGGGATGCTGGGGCCCCCATGCCCCCTCGCTGCCCTGCCA 1188
QY	829	T-ATTACATGATC-----ACCG-----ACTGAG-----AATATTGTTT--TA 863
Db	1189	TCCTTCTCTGTTCTCTTCTGCGGGGACACAGACACTGGGGCTCACCTCTTGGTGTATC 1248
QY	864	CAATAGTCTGTGGGCTGT-TTTTTTGTATTAACAAATAATTTAGATGTTGTAATAA 921
Db	1249	CTCTTGTACTG-GGAGAGGTGCTTTTGT--ATCCCCAATTA-----AAGGTAGAAA 1298

PR	26-AUG-1998;	98US-0097954P.
PR	26-AUG-1998;	98US-0097955P.
PR	26-AUG-1998;	98US-0097971P.
PR	26-AUG-1998;	98US-0097974P.
PR	26-AUG-1998;	98US-0098014P.
PR	01-SEP-1998;	98US-0098716P.
PR	01-SEP-1998;	98US-0098723P.
PR	02-SEP-1998;	98US-0098803P.
PR	02-SEP-1998;	98US-0098821P.
PR	02-SEP-1998;	98US-0098843P.
PR	09-SEP-1998;	98US-0099602P.
PR	10-SEP-1998;	98US-0099741P.
PR	10-SEP-1998;	98US-0099754P.
PR	10-SEP-1998;	98US-0099763P.
PR	10-SEP-1998;	98US-0099812P.
PR	15-SEP-1998;	98US-0100388P.
PR	16-SEP-1998;	98US-0100662P.
PR	16-SEP-1998;	98US-0100664P.
PR	16-SEP-1998;	98US-0101751P.
PR	16-SEP-1998;	98US-0101933P.
PR	17-SEP-1998;	98US-0100683P.
PR	17-SEP-1998;	98US-0100684P.
PR	17-SEP-1998;	98US-0100919P.
PR	17-SEP-1998;	98US-0100930P.
PR	18-SEP-1998;	98US-0100849P.
PR	18-SEP-1998;	98US-0101014P.
PR	23-SEP-1998;	98US-0101068P.
PR	23-SEP-1998;	98US-0101471P.
PR	23-SEP-1998;	98US-0101472P.
PR	23-SEP-1998;	98US-010475P.
PR	23-SEP-1998;	98US-010477P.
PR	24-SEP-1998;	98US-0101738P.
PR	24-SEP-1998;	98US-0101743P.
PR	24-SEP-1998;	98US-0101922P.
PR	25-SEP-1998;	98US-0101788P.
PR	29-SEP-1998;	98US-0102207P.
PR	29-SEP-1998;	98US-0102240P.
PR	29-SEP-1998;	98US-0102330P.
PR	30-SEP-1998;	98US-0102487P.
PR	30-SEP-1998;	98US-0102570P.
PR	30-SEP-1998;	98US-0102571P.
PR	01-OCT-1998;	98US-0102684P.
PR	01-OCT-1998;	98US-0102687P.
PR	02-OCT-1998;	98US-0102965P.
PR	06-OCT-1998;	98US-0103258P.
PR	06-OCT-1998;	98US-0103449P.
PR	07-OCT-1998;	98US-0016897B.

Query Match 43.2%; Score 2491; DB 6; Length 1300;

Best Local Similarity 42.6%; Pred. No. 3.5e-104;

Matches 562; Conservative 0; Mismatches 299; Indels 458; Gaps 83;

QY	1	GGGGCCCATCTGTTTCAGC-----CAGTCCCAAG---AATCATGAAAGTCGCCAGTG 51
DB	40	GTGGGAGCGCGGTAGCAGCAATGGACTTTCTCTGGGNAACCGTTTCAGCTCTCCAGTG 99
QY	52	G--CAGCAC--CGCCACCOCGCGCGGC--CCAGCTGCGCGCTGA--AGCC----- 98
DB	100	GGACAGCGCATCGAAGAACCCACAGATGGCTCCCTCCAGAGCGAGGACTGGGCCCTCAAC 159
QY	99	--GG-----CAAGACAG-----CGAGCG 114
DB	160	ATGAGATCTGGACATCATCAACGACGAGGAGGTCCCAAGATGCCCTCCGACGA 219
QY	115	GT-----CCGG-----CGAGTGGTGC--GCTGT--CTGTC 142
DB	220	GTAAGAGAGAGATCTGGGGATAGAACTTCCACAGGTGATGCTGGCTCTCACAGTC 279
QY	143	TGAGCA-----GAGCGTGGCCATCTCGCGT-----GCCGGGCG 176
DB	280	TTAGAAACCTGTGTCAAGAACTCGCGGCACCGCTTCCACGTGCTGGTGGCCAGCAGAC 339

QY	177	GCGGGGGCGG-----CCTGCTGCCCTGCTGGACGAGCAGCAGGTAACG 222
DB	340	TTCTGGAGAGTGTCTGTGTAGGACCATCTGCCCC--AAGAACACCCACCCACCATCG 397
QY	223	TGC-----TGCTCTACGACAT---GAAC--GGCT-----GTTACTCA---CGCC-- 258
DB	398	TGCATGACAAAGTGTCTAACCTCATCCAGTCTGGGCTGACGCGTTCCGACCTGCCCG 457
QY	259	-TC---AAGGACTGGTGCCCA-----CCCTGC-----CCC--AGAACCGC 293
DB	458	ATCTGACAGGTG--TGTCTACCATCTATAGGACCTGCGGAGGAAAGGCGCTGGAGTCCCC 516
QY	294	AAGG-----TGAGCAAGGTGGAGATTCTCCA-----GCACGTCTACGACT--- 333
DB	517	ATGACTGACCTGGACATGCTGTCAACCATCCACACCCAGAGGACCGTGTTCACACTCAG 576
QY	334	----ACATCAGGAC---CTTCAGTTGGAGCTGAATCGGAATCGAAGTTGGGACCCCC 386
DB	577	AGACACAATCAGGACAGGATCTGTGGCACTG--ACTC--CAGCCAGCAAGAGGACTCT- 632
QY	387	GGGGGC--GAGGCTGCCGCTCGGCTCAGCAC--CCTCAAAG---CGAGATC 440
DB	633	---GGCCAGCATGCTGCCCTCTG---CCGCGCCGCCCATACTCTCCGGTGACAGCC 686
QY	441	AGCGCCTCAAGCGCGAGCGGCGATCGGTTCTGGG-----GACGATCGC----- 485
DB	687	ATAGCACCAAC--CCCGAACAGATTGGGAAGCTGCCAGTGAGTGGAGATGGTGAAGTG 745
QY	486	-ATCTTG-----TGTC-----GCTGA---AGC---GCCTCCCCCAG---GGACCG 521
DB	746	GAACGTAGGGTATGTCGAGATGCTGACGAGTGGTGCC--CACCCAGCGGAGCCCG 804
QY	522	CGGACC-----CC-AGCCATCCAG---GGGGCA 545
DB	805	CAGACCTGGAGCTGTGAGGAGCTCAACCGCACGTCGCCGAGCCATGACGAGCGGGTCC 864
QY	546	AGAGGAATTACGTCTCTGGGTG--TCCCCACAGCGCC--TC---GCCGATCTGAG 598
DB	865	TGAGTGATACCTGTCTC--CGGGCCCATGCCCAAGAGCCCTTCAGAGCCACACTGCC 922
QY	599	GGAGAACAAAGCCGATC---GGC--GGCCACTG-----CG-----CCC 631
DB	923	AG---TCGAGGCTGCTGGAGGCTGCCACAGTGGAAATTCGCCGAGCCTATTGTCCC 979
QY	632	TTAACTGCATC--CAGCCTGGG--CTGAGGCTGAGGCACTGGC----- 671
DB	980	TACCTCTG--TCTGCTGATGGGCCCCCATGGCTTTGG--CTGGCCACTGAGGTAGGGTG 1036
QY	672	--CAGGAGAGGGGCTCTCTCTGCACACTACTAGTC--ACCAGAGACTTTAGGGGGTG 727
DB	1037	TGAGGTTGTGAGGCCCC--CTGAGGAGTGC--GGCGGCCACAGTAC---GAAGCTG 1087
QY	728	GGATTCCACTCG--TGTC-----TTTCTATTTTGAAGAGCAGACATTTTAAATAATG 781
DB	1088	CAACTCTGCGCGAGTGGCGAGATCTCAT-----CAGCC-----CCAGGC 1128
QY	782	CAGTTTGGTCTTCTCAGATTCTCAGG-----AAATTTGTTTATTATTGTA 828
DB	1129	TGAGGTGAGCTTTCAGGAGTCTGGGCCCCACTGCCCTCCGCTGCTGCTCCCTCCA 1188
QY	829	T-ATTACAATGATC-----ACCG-----ACTGAG-----AATTTGTTT--TA 863
DB	1189	TCCTTCCTCTGTTCTTCTTGGCCGCAACACAGCACTGGGCTCACCTCTTGGTTGATC 1248
QY	864	CAATAGTCTGTGGGCTGT-TTTTTTTTATTAAACAAATATTTAGATGGTGAATAA 921
DB	1249	CTCTGTACTG--GGAGAGTGCCTTTGT--ATCCCAATTA-----AAGGTAGAAA 1298

RESULT 9
ABU96435
ID ABU96435 standard; protein; 1300 AA.

KW tumour necrosis factor-alpha release; TNF-alpha release;
KW chondrocyte proliferation; chondrocyte differentiation; tumour;
KW adrenal tumour; lung tumour; colon tumour; breast tumour;
KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
XX
OS Homo sapiens.
XX
XX US2003032127-A1.
XX PD 13-FEB-2003.
XX PF 26-JUN-2002; 2002US-00183012.
XX 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 21-NOV-1997; 97US-0066120P.
PR 24-NOV-1997; 97US-0066466P.
PR 24-NOV-1997; 97US-0066772P.
PR 11-DEC-1997; 97US-0069335P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078886P.
PR 20-MAR-1998; 98US-0078939P.
PR 27-MAR-1998; 98US-0079664P.
PR 27-MAR-1998; 98US-0079786P.
PR 31-MAR-1998; 98US-0080107P.
PR 31-MAR-1998; 98US-0080194P.
PR 01-APR-1998; 98US-0080327P.
PR 01-APR-1998; 98US-0080333P.
PR 08-APR-1998; 98US-0081049P.
PR 08-APR-1998; 98US-0081070P.
PR 08-APR-1998; 98US-0081195P.
PR 15-APR-1998; 98US-0081838P.
PR 21-APR-1998; 98US-0082568P.
PR 21-APR-1998; 98US-0082569P.
PR 22-APR-1998; 98US-0082704P.
PR 22-APR-1998; 98US-0082797P.
PR 28-APR-1998; 98US-0083322P.
PR 29-APR-1998; 98US-0083495P.
PR 29-APR-1998; 98US-0083496P.
PR 29-APR-1998; 98US-0083499P.
PR 29-APR-1998; 98US-0083559P.
PR 06-MAY-1998; 98US-0084366P.
PR 06-MAY-1998; 98US-0084414P.
PR 07-MAY-1998; 98US-0084640P.
PR 07-MAY-1998; 98US-0084643P.
PR 07-MAY-1998; 98US-0085579P.
PR 15-MAY-1998; 98US-0085580P.
PR 15-MAY-1998; 98US-0085582P.
PR 15-MAY-1998; 98US-0085700P.
PR 18-MAY-1998; 98US-0086023P.
PR 22-MAY-1998; 98US-0086392P.
PR 22-MAY-1998; 98US-0086486P.
PR 28-MAY-1998; 98US-0087038P.
PR 28-MAY-1998; 98US-0087208P.
PR 02-JUN-1998; 98US-0087609P.
PR 02-JUN-1998; 98US-0087759P.
PR 03-JUN-1998; 98US-0087827P.
PR 04-JUN-1998; 98US-0088025P.
PR 04-JUN-1998; 98US-0088028P.
PR 04-JUN-1998; 98US-0088029P.
PR 04-JUN-1998; 98US-0088033P.
PR 04-JUN-1998; 98US-0088326P.
PR 05-JUN-1998; 98US-0088167P.
PR 05-JUN-1998; 98US-0088202P.
PR 05-JUN-1998; 98US-0088212P.
PR 05-JUN-1998; 98US-0088217P.
PR 09-JUN-1998; 98US-0088655P.
PR 10-JUN-1998; 98US-0088722P.
PR 10-JUN-1998; 98US-0088738P.
PR 10-JUN-1998; 98US-0088740P.
PR 10-JUN-1998; 98US-0088811P.
PR 10-JUN-1998; 98US-0088824P.
PR 10-JUN-1998; 98US-0088825P.
PR 10-JUN-1998; 98US-0088826P.
PR 11-JUN-1998; 98US-0088861P.
PR 11-JUN-1998; 98US-0088863P.
PR 11-JUN-1998; 98US-0088876P.
PR 12-JUN-1998; 98US-0089090P.
PR 12-JUN-1998; 98US-0089105P.
PR 16-JUN-1998; 98US-0089512P.
PR 16-JUN-1998; 98US-0089514P.
PR 17-JUN-1998; 98US-0089538P.
PR 17-JUN-1998; 98US-0089598P.
PR 17-JUN-1998; 98US-0089653P.
PR 18-JUN-1998; 98US-0089908P.
PR 19-JUN-1998; 98US-0089952P.
PR 22-JUN-1998; 98US-0090246P.
PR 22-JUN-1998; 98US-0090252P.
PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090429P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090461P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090688P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095598P.
PR 10-AUG-1998; 98US-0096012P.
PR 10-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097954P.
PR 26-AUG-1998; 98US-0097955P.

890 GTTA---TTAAACAATATTTAGATGTTGTAATAAAAAA 926
 1517 GCGAGCGCAAGCCAGCAAACTCATTCGGGAGCGAGGAA 1556

RESULT 5
 AAE37007
 ID AAE37007 standard; protein; 1811 AA.
 XX AAE37007;
 AC AC
 XX XX
 DT 23-OCT-2003 (revised)
 DT 07-AUG-2003 (first entry)
 XX XX
 DE Micromonospora carbonacea polyketide synthase (PKS) type I #13.
 XX XX
 KW Macrolide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme.
 XX OS Micromonospora carbonacea.
 XX XX
 FH Key Location/Qualifiers
 FT Misc-difference #52..1811
 FT /note= "Encoded by GCC"
 XX CA2391131-A1.
 XX 19-NOV-2002.
 PD 26-JUL-2002; 2002CA-02391131.
 PP 26-JUL-2001; 2001US-0307629P.
 PR (ECOP-) ECOPIA BIOSCIENCES INC.
 PA Yang X, Staffa A, Farnet CM;
 XX WPI; 2003-343556/33.
 DR N-PSDB; AAD55823.
 XX Novel isolated polypeptide involved in biosynthesis of macrolides by
 PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
 PT preferably for biosynthesis of rosaramicin.
 XX Claim 13; Page 183-189; 206pp; English.
 CC The invention relates to genes and proteins involved in the biosynthesis
 CC of macrolides by microorganisms. In particular it relates to the nucleic
 CC acids forming the biosynthetic locus for rosaramicin (a 16-member
 CC macrolide antibiotic) from Micromonospora carbonacea. The invention is
 CC useful for the biosynthesis of macrolides by microorganisms. It allows
 CC direct manipulation of macrolides and related chemical structures by
 CC chemical engineering of the proteins involved in the biosynthesis of
 CC rosaramicin. It is useful to catalyse certain biochemical reactions, in
 CC vitro or in vivo, to direct or enhance the synthesis or modification of a
 CC polyketide, polyketide substrate or its precursor. The present sequence
 CC is M. carbonacea polyketide synthase (PKS) type I protein. (Updated on 23
 CC -OCT-2003 to standardise OS field)
 XX Sequence 1811 AA;

Query Match 44.4%; Score 2562; DB 6; Length 1811;
 Best Local Similarity 41.4%; Pred. No. 3e-107;
 Matches 572; Conservative 0; Mismatches 313; Indels 498; Gaps 86;

QY 5 CCCATTCTGTTTCAGCCAGTCGCCAAGATC-----ATGAAGTCGC-CAG-----TG 51
 DB 465 CCCAGTGTACCGGATCTCGACGCGATCTCGCCCGAGAGACCGAGCGCGGAGCTG 524
 QY 52 CGACACCGCC-----ACCGCGCGCG--CGGCGCCAGCTCGCGC-----TGAAG-- 96
 DB 525 CGCGCCCTGCGGTGCGCTCTCTCTACCGGCGCGGTGACGCGTGCACAGAGGACGACGCGG 584
 QY 97 -----CGGCGAGA-----CAGCG--AG-CGGTGGCGGC-GAGGTG 128

217 TAAACGTG---CTGC-----TCTAC---GACATGAACGGCTGTACTCAAGCC--- 258
 458 AGATCAAGAGCTGCTGAGATGTTCAAGTGGACATGAACGACATGTAGCCCTACCTCC 517
 259 ---TCAAGAGCTGTGTCACAC-CTGCGCCAGAACCG-CAAG-GTGAGC-AAGGTGGAG 311
 518 TCTTCAAGGAGT--GTGACCACTCCAAAC--GACCGTCTAGAGGGGGCTGAGATCGAG 573
 312 ATTCTCCAGC-ACGTC-----ATCGACTACATCAGG---GACCTTCAGTTGAGCTGA 360
 574 GAGTTCCTCGCGGGCTGCTGAAGCGCGCGAGCTGGAGGAGATCTTCAT-----CAGT 628
 361 ACTCGG-----AATCCG-AAGTTGGACCCCGGGG--GCCAGGG--CTGCCGTCC--- 408
 629 ACTCGGGCGAGACCGCGTCTGAGTGGCCCTGAGCTGCTGAGTTCCTGAGGACACAGG 688
 409 -----GGGCTCGCTCAG---CACCTCAACGCGAGATCAG-----CGCCC 447
 689 GCGAGAGGGCGGACACTTGGCGCGGCGCCAGCAGCTCATTCAGACCTATGAGTCAAG 748
 448 TGACGCGCGAGCGGCGATCGTTCCTGCGGAC---GATCGCATTTG-----TG 493
 749 AGACAGCC-AAGCAGCATGAG---CTGATGACACTGGATGGCTTCATGATGATGATGTTG 804
 494 TCGCTGAAG-----GGCTCCG---CCAG-----GGACCGG 521
 805 TCGCCGAGGGGAGCTGCCCTTGGACAAACCCACACAGTGTGTTCAGGACATGACACAG 864
 522 C-----GGACCCAGC--CATC-----GGCTCCG-----C 537
 865 CCCCTTGGCCACTTCTCTCTCTCCACACACCTATCTGACTGACTCCAGATC 924
 538 AGGGGGAGAGGAATTAAGTG--CTCTGT--GGGTCT--CCC-----CCACGCG- 582
 925 GGGGGGGCCC-AGCAGACCGAGGCTTATGTTAGGCGCTTGGCCAGGATGCGGTGCT 983
 583 -----GGATCT-----GAGGAGAAACAGACCAATCGGGGCA----- 623
 1044 CACCTCCAGATTTCTCTTCGGGACGTGGCCCAAG-CCGTGC-GGACCATCGCTTCAG 1101
 624 CTGCGCCCTTA-ACTG-CATCCAG--CCTGGGG-----CTG-AGGCTGAGCA-CTGGCG 672
 1102 CTGTCCCTTACCTGTCTATCTATCCCTGGAGAACCACTGGGGCTGGAGCAGCAGG- 1160
 673 AGGAGAGGGCGCTCCT-CTCTGCAC-----ACCTACTAG----- 705
 1161 -TGCCATGGCCCGCCACCTCTGCACCATCTGGGGGACATGCTGTGACACAGCGCTGG 1219
 706 -----TCACCAGAG--ACTTTAGGG-----GGT-- 726
 1220 ACTCCCAAAATCCGAGGAGCTGCGATCCCGACAGCAGTGAAGGCGCGGTCTGTGTA 1279
 727 -GGGATT-----CCACT-----CGTGT-----TCTAT----- 750
 1280 AGGGAAGAAGCTGCCCGTCTCGGAGCGAGATGGCCGGGCTCTGTGCGATCGGGAG 1339
 751 -----TTTTTGAACGACACATTTTAAATATGTT-----CA-----CGTT 787
 1340 AGGGGAGGAGGATGACGAGGAGGAAGAGAGAGGTGAGGCTGAGGCGCAGGCGCG 1399
 788 TGTGCTTCTCAGATTTCTGAGGAAT---TGCTTTGTATTGTATTATTAATATGATCACC 844
 1400 TG--GCCAAGCAGATCTCCCGGAGCTGTGCGGCCCTG-GCTGTGTACTGCGCACGCC 1456
 845 GACTGAGATATTGTTT-----TACATAGTTCTG-TGGGCTG-TTTTTT 889
 1457 GCCTGCGGACCTGACCTGACCTGCGCCCAACGCCCCCAACCCCTGCCAGGTAGCTCCCTCA 1516

CC glomerulonephritis, Hashimoto's thyroiditis, ischaemic injury, multiple
CC myeloma, multiple sclerosis, osteoporosis, Parkinson's disease, psoriasis
CC and vasculitis. Disorders may be treated with the novel composition using
CC gene therapy. This polynucleotide sequence represents a TALL-1 related
CC DNA sequence of the invention
XX
SQ Sequence 7285 AA;
Query Match 46.5%; Score 2691.5; DB 6; Length 7285;
Best Local Similarity 37.1%; Pred. No. 3.9e-112;
Matches 620; Conservative 0; Mismatches 294; Indels 757; Gaps 83;
1 GGGGCCCATCTTCTGTTGAGCCAGTCG---CCAGAA----- 33
Db 2671 GATCCCGCGTGGAGGATCATCAGCCGGCGTCCGGAACGATTCGAGCCCAACC 2730
Qy 34 --TCATGAAGTCCGAGTGGACACCGCCAC-----CGCGC----- 70
Db 2731 TTTTCATAGAGCGCGGTGGAATCGAAATCTCGTATGCGAGTGTGGCGTCTGTTGT 2790
Qy 71 CG-----CGGGCCCGAGTGGCGCT--GAAGGC--CGGCAAGACACGA-----GC 113
Db 2791 CGGTCTNTTCGAACCCAGAGTCCGCTCAGAGAACTGTCAAGAGCGATAGAGGC 2850
Qy 114 GGTG-----CGGGCGAGTGGTGC-----GCTGTCTG-----TC 142
Db 2851 GATCGCTCGAATCGGGAGCGCGATACCGTAAAGACGAGGAGCGTCAAGCCATTC 2910
Qy 143 TGACGAGC-----GTGGCCATCTC-----GCG----- 166
Db 2911 GCGGCCAAGCTTTCAGCAATATACGGGTAGCGAAGCTATGCTCTGTAGCGTCCG 2970
Qy 167 -----CTGCGCG----- 173
Db 2971 CACACCCAGCGCGCACAGTCAATGAATCCAGAAAGCGCCATTTTCCACCATGATTT 3030
Qy 174 -----GGCGCGCGG-----GGCGCGCT- 190
Db 3031 CGGCAAGCAGCATCGCATGATGATCAGACAGATCCTCGCTCGGGCATGCGGCTT 3090
Qy 191 --GCTGCGC-----CTGCTGG--ACGACAGCAGG--TAAACGTGTGCTC----- 230
Db 3091 GAGCTGCGCAACAGTTTGGCTGGCGGAGCGCCCTGATGCTCTCTCCAGATCATCTG 3150
Qy 231 TAGCATGTAACGGCT-----GT--TACTCAGC-----CT----- 259
Db 3151 ATCGACAGACCGGCTTCCATCCAGTACGCTGCTCGTCAAGCGATGTTTCTGTTGGT 3210
Qy 260 -----CAAGAGGTGG-----TGG-----CCACCCCTGCCCGAAGCC--- 291
Db 3211 GTCGAATGGCAGGTAGCGGATCAAGCGTATGACGCGCGCGCATTTGCATCAG--CCATG 3268
Qy 292 -----GCAAGGTGAGCAAGGTGGAGATCT--CCAGCAGCTCA 327
Db 3269 ATGNTACTTTCTCGGCAGGAGCAAGTGAAGTACAGAGAGATCCTGCCCGCGCTTGG 3328
Qy 328 TGGACTA--CATCAGGACCTTC-----AGTTGG-----AGCTGAATCGGA 367
Db 3329 CCAATAGACCGCATGCTTCCCGTTTCAAGTCAACAGTCAAGCAGTGGCAAGGA 3388
Qy 368 A--TCCGAAGTTGGACCCCGGGCGCGGAGGGTGC----- 402
Db 3389 AGCGCGGTGTTGGCCAGCAGCAGTAGCGCG--GCTGCTCGTCTGCAATTCATTCAGGA 3446
Qy 403 -----CGGT-----CCGGCTCGCTCAGC-----ACC-----CT 427
Db 3447 CACCGGACAGTGGTCTTGACAAAAAGAACCGGGGCGCCCTCGCTGACAGCGGAGCA 3506
Qy 428 CACCGCGG--AGATCAGC-----GCCCTGAC--GGCGG----- 456
Db 3507 CGCGCGCATCAGACAGCGGATTTGCTGTTGTGCGCCAGTCAAGCGCAATAGCCTCTCCA 3566
Qy 457 ---AGCGCGCATGCTTCTCTCGGACGATCGCATCTTGTGCTGCTGAAGCG-----CC 506

Db 3567 CCCAAGCGCGCGAGAACCTGGGTGCAATC-CATCTTGT-TCAATCATGCGAAACGATCC 3624
Qy 507 TCCCCCAG-----GGACCGGC---GGACCCC--AGCCATCC-----AGGGGGCAAG-A 548
Db 3625 TCATCTCTCTCTTGTATCTGATCTGATCCCTGCGCCATCAGATCTTGGGGGCAAGAA 3684
Qy 549 GGAAATACG--TGCTCTGTGGGTCTCCGCCAAC-----CGGCTCTG----- 587
Db 3685 AGCCATCCAGTTTACTTTTCAGGGCTTCCCAACCTTACCAGAGGGCGCCAGCTGGCAA 3744
Qy 588 --CCGATCTGAGGGAGAAC--AAGACCG-----ATCGGC-----GGCCACT 625
Db 3745 TTCGGTTCGCTGCTGCTCATAAACCGCCCACTAGCTATCCCATGTAAAGCCACT 3804
Qy 626 GC-----GCCCT-TAACT 637
Db 3805 GCAAGCTACTGTTTCTCTTTTGGCGTTTTCGTTTCCCTTGTCCAGATAGCCAGTAGCT 3864
Qy 638 G-CATCCAGCCTGGGGCTG-----AGGCT----- 660
Db 3865 GACATTCATCCGGGGTTCAGCACCGTTTCTGCGGACTGGCTTCTACGTTTCCGCTTCT 3924
Qy 661 ---GAGGCACTGGCAGGAGAGGCG-----GCTCC-- 687
Db 3925 TTAGAGCCCTTTCGCGCTGAGTGTGTCGCGCAGCGTGAAGCTACATATATGTATCGG 3984
Qy 688 -----TCTCTGCA-----CACCTAC--TAGTCAACAGAGAC-----TTT 719
Db 3985 GCAATCGCTGAATATTCCTTTTGTCTCGCACCATCAGGACCTGAGTCTGCTGCTTTT 4044
Qy 720 AGGGGTGGGATTC-----CACTC-----GTG-----TGTTT 746
Db 4045 C-----GTGACATTCAGTTCGCTCGCTCACGCTCTGGCAGTGAATGGGGTAAATGGCA 4100
Qy 747 CTA-----TTTTTTGAA--AAGCAG-----ACATTTTAAAAAATGGTCACT 786
Db 4101 CTACAGCGGCTTTTATGATTCATCAAGAGAACTACCCATAATACAAGAAAGCCCGT 4160
Qy 787 TTGCTGTTCTCA---GATTTCTGAGGAAATGCTTTGTATTG-TATATTAC-----AA 836
Db 4161 CACGGGCTTCTCAGGGCGTTTATGGCGGCTGCTATGTGGTGTATCTGACTTTTTC 4220
Qy 837 TGATCA-----CCGACTGAGAAATATTTTCTTAC-----AATAGTTCTGTG--- 876
Db 4221 TGTTCAGCAGTTCTCGCTGCTGATTTTCCAGTCTGACCACTTGGATTATCCCGTGACA 4280
Qy 877 GGGCTGTTTTTGG--TTATTAAACAATAATTTAGATGGTGAAGAAAAAA 925
Db 4281 GGTCAATTCAGCTGGCTAATGACCCAGTAAGGAGC-GGTATCATCAACA 4330

RESULT 2

ABG70822
ID ABG70822 standard; protein; 3907 AA.
XX
AC ABG70822;
XX
DT 16-DEC-2002 (first entry)
XX
DE Mouse myocardin associated protein #1.
XX
KW Mouse; myocardin; myocardial infarction; cardiomyocyte;
KW post-mitotic cell; differentiation; therapeutic; gene therapy;
KW heart disease; cardiomyopathy; mortality; heart failure; hypertension.
XX
OS Mus sp.
XX
FN W0200260946-A2.
XX
PD 08-AUG-2002.
XX
PF 21-DEC-2001; 2001WO-US050606.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:10:55 ; Search time 66 seconds
(without alignments)
3964.227 Million cell updates/sec

Title: X77956
Perfect score: 5766
Sequence: 1 GGGGCCCATCTGTTTCAGC.....TTAGATGCGTGAAGAAAAA 926

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003s:*
- 7: Geneseq2003ss:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2691.5	46.5	7285	6	ABJ38280 pAMG21-EA
2	2648	45.9	3907	5	ABG70822 Mouse myo
3	2648	45.9	3907	6	ABG74190 Mouse myo
4	2596	45.0	2211	6	ABG72959 Novel hum
5	2562	44.4	1811	6	AAE37007 Micromono
6	2517	43.7	1329	5	AAU91279 Human NOV
7	2491	43.2	1300	6	ABU88254 Novel hum
8	2491	43.2	1300	6	ABU90133 Novel hum
9	2491	43.2	1300	6	ABU96435 Novel hum
10	2491	43.2	1300	6	ABU99044 Novel hum
11	2491	43.2	1300	6	ABU98259 Novel hum
12	2491	43.2	1300	6	ABU91965 Novel hum
13	2491	43.2	1300	6	ABU95269 Novel hum
14	2491	43.2	1300	6	ABO00408 Novel hum
15	2491	43.2	1300	6	ABU88959 Novel hum
16	2491	43.2	1300	6	ABO06455 Novel hum
17	2491	43.2	1300	6	ABU95515 Novel hum
18	2491	43.2	1300	6	ABU95205 Novel hum
19	2491	43.2	1300	6	ABU90753 Novel hum
20	2491	43.2	1300	6	ABU93915 Novel hum
21	2491	43.2	1300	6	ABU86189 Novel hum
22	2491	43.2	1300	6	ABU82044 Novel hum
23	2491	43.2	1300	6	ABU07905 Novel hum
24	2491	43.2	1300	6	ABU94225 Novel hum
25	2491	43.2	1300	6	ABO00098 Novel hum

26	2491	43.2	1300	6	ABU87109	Novel hum
27	2491	43.2	1300	6	ABU91350	Novel hum
28	2491	43.2	1300	6	ABU90443	Novel hum
29	2491	43.2	1300	6	ABU97034	Novel hum
30	2491	43.2	1300	6	ABO05230	Novel hum
31	2486.5	43.1	1679	4	AAU07343	i-aminocy
32	2486	43.1	1089	6	ABM15834	Mycobacte
33	2445.5	42.4	1743	6	ABU88255	Novel hum
34	2445.5	42.4	1743	6	ABU90134	Novel hum
35	2445.5	42.4	1743	6	ABU96436	Novel hum
36	2445.5	42.4	1743	6	ABU99045	Novel hum
37	2445.5	42.4	1743	6	ABU98260	Novel hum
38	2445.5	42.4	1743	6	ABU91966	Novel hum
39	2445.5	42.4	1743	6	ABU85270	Novel hum
40	2445.5	42.4	1743	6	ABO00409	Novel hum
41	2445.5	42.4	1743	6	ABU88960	Novel hum
42	2445.5	42.4	1743	6	ABO06456	Novel hum
43	2445.5	42.4	1743	6	ABU95516	Novel hum
44	2445.5	42.4	1743	6	ABU95206	Novel hum
45	2445.5	42.4	1743	6	ABU90754	Novel hum

ALIGNMENTS

RESULT 1

ABJ38280
ID ABJ38280 standard; protein; 7285 AA.
XX
AC ABJ38280;
XX
XX
DT 12-JUN-2003 (first entry)
XX
DE PMG21-RANK-Fc vector protein SEQ ID NO 28.
XX

TALL-1-binding protein; TALL-1; B-cell-mediated autoimmune disease;
systemic lupus erythematosus; B-cell-mediated cancer; lymphoma;
inflammation; rheumatoid arthritis; acute pancreatitis; atherosclerosis;
Alzheimer's disease; asthma; cachexia; cirrhosis; diabetes; osteoporosis;
glomerulonephritis; Hashimoto's thyroiditis; ischaemic injury; psoriasis;
multiple myeloma; multiple sclerosis; Parkinson's disease; vasculitis;
gene therapy; ds.
XX
OS Unidentified.
XX
PN WO200292620-A2.
XX
PD 21-NOV-2002.
XX
PF 13-MAY-2002; 2002WO-US015273.
XX
PR 11-MAY-2001; 2001US-0290196P.
XX
PA (AMGE-) AMGEN INC.
XX
PI Min H, Heu H;
XX
PS WPI; 2003-156719/15.
XX

New TALL-1-binding polypeptide, useful for modulating the activity of
TALL-1 and in treating, preventing or diagnosing a B-cell-mediated
autoimmune diseases, cancers or lymphomas.
XX
PS Disclosure; Fig 4; 236pp; English.
XX
XX The invention relates to a novel TALL-1-binding polypeptide comprising a
defined sequence in the specification. The composition is useful in
modulating the activity of TALL-1, and in treating, preventing,
ameliorating, diagnosing or prognosing a B-cell-mediated autoimmune
disease (e.g. systemic lupus erythematosus) or B-cell-mediated cancer or
lymphoma. The composition may also be used in treating inflammations
CC (e.g. rheumatoid arthritis), acute pancreatitis, Alzheimer's disease,
asthma, atherosclerosis, cachexia, cirrhosis, diabetes, CC

2491	NTTAXCRLLT	CADIQNGT	ISTSV	CVV	ALST	CVS	NGT	ACI	PFK	KANC	STY	TK	AC	NS	G	G	L	D	I	2550
DB						:	:	:	:	:	:	:	:	:	:	:	:	:		
QY	1191	---	TTTT	TATA	AAA	AGT	1202													
DB	2551	CVFT	Q	T	A	T	A	G	A	G	A	G	T	2565						

Search completed: May 7, 2004, 15:10:02
Job time : 74 secs


```
RESULT 14
Q8VIZ1 PRELIMINARY; PRT; 1384 AA.
AC Q8VIZ1;
DT 01-WAR-2002 (TrEMBLrel. 20, Created)
DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PE_FGRS family protein.
GN MT3612.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]_TaxID=1773;
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., PeBooy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uutterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains." to the EMBL/GenBank/DBJ databases.
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007163; AKA47970.1; -.
DR TIGR; MT3612; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000084; PE region.
DR InterPro; IPR002173; PFKB.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.
DR PROSITE; PS00583; PFKB_KINASES_1; 2.
SQ SEQUENCE 1384 AA; 110839 MW; 124AC8AAADD5EBD CRC64;

Query Match 11.5%; Score 869.5; DB 16; Length 1384;
Best Local Similarity 30.4%; Pred. No. 2.2e-43;
Matches 418; Conservative 23; Mismatches 612; Indels 321; Gaps 54;

QY 1 GATCTGGGTGCTGCCAGGAAAGCAAAATTCGGAAGTTAAATGTTTTCAGTGAATTTT 60
DB 147 GAAGCGGNGSAGLWNGNGAGGAG-----GSGGAGGNGGNGWLFAGAGTGGTGGTG 200
QY 61 AAATCTCTCTCGCGGAGA---GGCCCGCTCTCCCGGATATCAGCGCTTCCATCTTT 117
DB 201 APG---AMGCTGGNGGALLIG-----GGLGAGGAGGCTGGTGGTGGNG 244
QY 118 TGAATCCGCGGCTCCGCGTCTTCGCGCTCA-GACCAAGCGGAGAGAGCTGTTTGA-- 174
DB 245 GNGALLIAGGVG--GAGG---TGGGTGAGAGAGGTGGNGAGGLEFWNGDGGAGG 298
QY 175 -----ATTAAAGCGG-----GCTGTGAAGCCCGAGCGCGCGCGCGCGCGGCC 218
DB 299 QGGDGAAGAAASAGTGGKGGQGGDGTGGAGAGPVLFGHGGAGGAGGAGTGGMG-- 356
QY 219 GAGCGCGGCAATT--TGAATAAGAGCGCTGCTTCAGGCGAGGCTCTATAAGTGACCG 276
DB 357 GAGGDTTVAAGTGGEGTGAAGAGG--AAGARGALTSGLAGG-----VGAGGTG--G 408
QY 277 CCGCGCGAGCGTGGCGCG--GTTGAGGTCACTAGCGGACCTCTTTTGGTTTCTTTC 335
DB 409 TGGTGGNGADAATAVVGFGANGDPGAGGK-----GNGGIGGA-----AVTGG-- 451
QY 336 TCTTTTGGGCACTCTTGACTCACTCCCGAGCATGAAGCGGCTGAGCCCGGTGCGCGGCT 395
DB 452 ---VAGDGTGGKGTGGA--GGAGNDAGSTNFGKGGDGGGAGGAGGAAGTGGGHA 507
QY 396 GCTACGA--GGCGGTGTGCTGCTG-----TTGGA-----ACCGAGTGGCCATCGCC 442
DB 508 GNTGCGDGGTGGNGGNGTGGVNGADNTLNPTDPCGAGPAGGAGGAGGAGGPGGTGG-- 565
QY 443 CGGGGCGGAGGGAAGGG-----CCCGGAGCTGAGGAGCGCGCTGAGCT-- 485
```

```
DB 566 TGNNGNGNGNGNGNGNGNGNGNGNAGNNSTNAPVGGEGAGGDDGAGAGGAANGGTAG 625
QY 486 ---TGCTGGACGACATGAACCACTGCT-----ACTCCCGCTTGGCGGAACGTGATC 533
DB 626 SQGTGGVGGDGGAGNGGGGKAGTNGSNGFVGDGAGFGSGAGNGGVGAAGANGGTGG 685
QY 534 CCGGAGTCCCG--AGAGGCACTCAGCTTAGCCA-----GGTGAATCTTACAGCGC 583
DB 686 SGNNGDGGAGGAGGAGGNGI PGTGTEPAGGTGAKGDDGDDGAGGAGNAGGAGNGGA 745
QY 584 ---GTCATCAGCTACATTCT---CGACTCGAGTAGTCTCTGCCCGGAGCAGCCCTG 635
DB 746 GGQGENAGCGGAGGAGGAGNAPGDGVGKAPHGDPAGSGGDDGKGGGSGGTGGSCAPITG 805
QY 636 GACCCCTGATGCGCCGCCACCTTCCCATCCAGACAGCCGAGCTCGCTCCGGAACCTTGA 695
DB 806 GG-----AGGTGG-----SGHAGKGGAGGI-----GAGGTIIVP 836
QY 696 TCTCCAAACGACAAAGAGG-----TTTTGCCACTGACTCGGCGGTCTCTGACACCTCC 750
DB 837 GNGNAGDGGNAGAGGNGGSGDFGNTTSG---ASGSGGNGNAGT-----A 880
QY 751 AGAACGCAAGTGTCTGGCGCCGCTTCTGCTGGACCCCG-----GGAACCTC 797
DB 881 GSGGAGGTGTGLSGGNGGNGGNGGNGGNGAGHGTGVAQFVPATSLPTPNAGGAGNGG 940
QY 798 TCTGCGCGGAAAGCGGACGCGAG-----GGATGGGCCCC-----AAC 834
DB 941 TGSNGGAPGAPGAPGPTTGGNAGSOGIGDGGNGSGDGGKGGDADAVNVVMPFPAAT 1000
QY 835 TTCGCCCTGCCCACTTGACTTACCAA-----GATGGGCCCC-----ATCCCTTCCTG 872
DB 1001 GTAG--SAGDPTGGNGGPGTGPSPWVAPPPTITVQOQGGDGGAGGTGTWANDGTATG 1058
QY 873 GAGACTAACTGTGTCTCAGGAGCGAAG--ACTGTGAATTGT-----AGCC 919
DB 1059 KGK-----EGVGSILGGPGNGGTGGNATGATGTVANAGNGKGGDGGGQFGAGN 1111
QY 920 TGAAGACCCAGAGTACTCTGCCC--ACCACTGGGCGACGCTCACCTGTCTCCACCCC 977
DB 1112 GGAGSVTDSAGSTAGNGGNGNATNGTIAQAPAGNGSAGKG--GPGGN-----1161
QY 978 ACCCCCAAGTCTAA--GGTCTTTTACAGCGGTGGAGGTGTGAGAGAGTGGCTCTCTCC 1036
DB 1162 -----IAAGATGTAGNGGNGNGNDGAVNAGTGGSGNG--GNAGGGGANGGGG-----1209
QY 1037 AAATATGCCAAGCGCGCGCAGAGCTGTCTTCTGCTCTCTCTTGGAGAAAGTCTCTGT 1096
DB 1210 -----AGGAGGAGGKGGIDGGFGDGGGNGNGTGAAGNGGNGGTGGVGSV 1258
QY 1097 GCCCTGATTATGAATCTATATAGATATAGTGTGTTTGTACCTTTTACAGGAAGG 1156
DB 1259 GAAGDGGNGGTGGPAGFGGTGNGGSG--GTGAGDGGGTGGDGGNGGTGTAGGGTGG 1317
QY 1157 TGACTTT--CTGTAACA-----ATGCGATGAT--ATTAACTTTTATATAAAGT 1202
DB 1318 NGGASGAGGAGGTGGFAGNGNAGNGGTGGASBDGNDGNAGSGATCGTGGNGGT 1371

RESULT 15
Q94710 PRELIMINARY; PRT; 2717 AA.
AC Q94710;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE S1A surface protein.
OS Paramesidium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Penicillida;
OC Paramesidium.
OX NCBI_TaxID=5889;
RN [1]
```

Db 431 TTCAAITLEAACVXGSGTGFWDTPASSCKDKICVNAPATNTTHDLQCFALNTCTVNST 490
Qy 162 AAGC-----CTGTTTGA-----ATTTAAGC---GGGCT- 187
Db 491 SAGCVKTCENSLVLAICDKDTSNRACIWKHKYKQCVLASSATTTTHADQTHSSCTL 550
Qy 188 ---GTG---AACGCCAGGGCGGGCGGGCGGGCGCGAG-----GC 223
Db 551 SNQGTGCVPLPLKCEAITLEAACN-MKANGQPCG-WNGSQCIDKACSTASKTATTTQC 607
Qy 224 CG-----GC----- 227
Db 608 QGHISTCVANNPATVNGSLTIQGCODLPTTCAARKSTENCEISRAQFPFCLWVASSSTCV 667
Qy 228 ---CATTTTGAATAAGAGCGGTGC-----TTCCAGGAGGC-----TCTAT 267
Db 668 EXSCATAGTIGALSAGGFTFSACQAVLNTICISNNADGCMWKPSSCSLVSNNCQLG 727
Qy 268 AAGTGAC-----C-----GC-----CGCGCGGAGC-----GTGC----- 291
Db 728 SKASGDCVWNGSSCDRTCANITQTTHNSCOSTFNQCTVYNGGTACQPLATACTSYSTQE 787
Qy 292 GC-----GCGTTG---CAGGTCA-----CT-----GTAGC- 313
Db 788 NCKFTSTNKNVWGLACRNATCADAPDTASVDTTDECLAYFPBPSETCTVYKVGQAQCV 847
Qy 314 -----GACTT----- 319
Db 848 PKSANCNVYMTQAQCHKITINLTANDCKWVDRCVAISSFATGACTTFFKGTQTMCOGYR 907
Qy 320 ---CTTTTGGTTTT-CTTCTCTT-TG-----GGGAC-CTCTGGAC 355
Db 908 AGCNTGGATSSFACTLDCTLTGTGLTFADCOALDSTCVSKDGTGCIVIOSACTGYGT 967
Qy 356 TCACTCCCAGCATGAAGCGCTGAGCC-----CG----- 385
Db 968 TAA---NCRSSASGNAGYCALNGSSTCVTVSAAADCAFTVGLTGLDHTKCOLYHTSCTA 1024
Qy 386 ---GTGC-----GCGGTGTGTAGAGCGGTGTGC-----TGCTGTGGA 423
Db 1025 LKDGTCQBYKTCGSAATNNCATSGQKCYPDGTDCLRFTNCASVTGTGLTTICATY 1084
Qy 424 ACCAGTCTGGCC---ATCGCCCGGGCGGAGGAGGCGCGG-----C 465
Db 1085 DAGCIANVDGTAQCKLAICDLVLTQKSCSTSAATAADKCAWSGTACLAIVTANIAIHC 1144
Qy 466 AGCTGAG-GAGCCGCTGAGTTCTGGAC-----GACATGAAC-----CACT 506
Db 1145 AYVTGTGLTDAICAAYNVDCTANRAGTACQEQKATCDLYTTEATCSTSAATAADKCAWS 1204
Qy 507 GCTACTCCCGCTGC---GGG-----AACT---GGTAC-----C 534
Db 1205 G-TACLAVTVAITHCAVKGTLTDTTCAAYNADCTANRAGTECQEQKATCDLYTTEATC 1263
Qy 535 -----CGGAGTCCGAGAGGCACTCA-----GCTTAGC-----CAGGTGG--- 569
Db 1264 STSKAATAADKCAWSGTACLAIAETANIAITHCAVKGTLTDTTCAAYNVDCTANRAGTEC 1323
Qy 570 ---AAATCTACAGCGGTCA---TCGACTACATTC-----TCGA 603
Db 1324 QEQKATCDLYTTEATCSTSAATAADKCAWSGTACLAIVTVAITHCAVKGTLTDTTCAA 1383
Qy 604 CCTCAGGTAGTGC-----TGGCGG---AGCCAGCCCTGGAC-----CC 640
Db 1384 YNADCTANRAGTECQEQKATCDLYTTEATCSTSKAAATADKCAWSGTACLAIVTVAITHCA 1443
Qy 641 CCTGATGCGCCCACTTCCG---ATCCAGACAGC-----CGAGTCTGCTC----- 683
Db 1444 YVTG-TGLVAATCAAYNVGINLKDGTGCGQEAQAGCKDYSTONKCSAQSTGSLUSCIWDS 1502
Qy 684 -----CG-----GAACTT-----GTCACTCT--- 698

Db 1503 SCYTLGEKNCSITGTGLDHAQCOQSYSGAGTSITDGTCKODLKSCEQYAGTTASCTKTA 1562
Qy 699 --CAAACGA-----CAAAGGAGCTT-----TTGC----- 721
Db 1563 TSKYLGSGPTCITFISNVATDCAKITGGAGTVTYDIQCSYNTGCSANRARSACVQQAQCS 1622
Qy 722 -----CACTGACTCGCGGTCTCTGACACCTC-----CAGAACGA-- 758
Db 1623 GYTVLTSYKSGAGLCIASNTTDTACVAASSATTCDAVLTGTYGNYNANCVEMKAGCTNN 1682
Qy 759 GGTGTC-----TGGCGC-----CCGTCTCTGC----- 778
Db 1683 GTTACVAKSCANAAGITFNHANCNSYLTCTVNSGGTACQTMIPKCADOTSTACLYSVEG 1742
Qy 779 -CTGGGAGC-----CCGGAACCT----- 796
Db 1743 ECVVIGTSCVRKCTDAIDAIDRDDDBSCTVYQOCTVARLGACQARAACGTYKSLQCK 1802
Qy 797 -----C-----TCCTGCCG-----GAA--GCC 811
Db 1803 FNTSGGKCFWNPNTKICVDLNCNIESTSFDTNHCVAVDSTLACTVRAINGAAVQGM 1862
Qy 812 G-GAC-----GGC----- 818
Db 1863 ARGACASYTIEQCKTNASNGICVMTNANLPAPAQDKSCTSAPTSTTTTHNDYAYNT 1922
Qy 819 -----AGGATCGGCCCAACTT-----CG-----CCCTGC- 844
Db 1923 ATVKCTVATPSNSGKNPTLGGCQQTAACSSYIDKECQINANGDPGWNGTQCADRACS 1982
Qy 845 -----CCACTTGACT-----TC-----ACCAAT 863
Db 1983 TAPATTDYDDTCKRAYLSNKTIVSDIGEGVEIPATCETMTQKCYSNKAGDPYWTGT 2042
Qy 864 CCCTTCTCTGAGA-CTAAACCTGGTCT-----C-----AGGAGCGAGGACTGT 907
Db 2043 TCTKSCDNAPDATATADECNTYLAGCTLDNVKCKTKVCDFAPATDALCKQALSTCTTN 2102
Qy 908 GAACTT-GT-----AGCCTGAAGAGC-----CAGAGC-TAGTCT 940
Db 2103 GTNCVTRGTFOALSQAGCVSSTTGOCEWIPAVLNASNVVTTTPAYCTIKNCSTAPITLT 2162
Qy 941 GGCCACAG---CT---GGCGACGTCACCTGCTCCACCCACC---C 981
Db 2163 SE--AACAGVFTNCTXNGGCVTKSTCAAVIDVACTTALNGTICAWDSAQNKCRDKDC 2220
Qy 982 CCAAGTT-----CTA-AGGTC----- 996
Db 2221 QDFSGTTHAACQAEAGCTAGASGKCAVQNCETQVRSACIEGTNGPCLWIDKYVNTDG 2280
Qy 997 -----TTTTCAAG--CGTGA-----GGTGGAGGAGGTGC- 1028
Db 2281 SKGACFRYTSWKSIAWNSDTSCKWISNKCTNGSNVCGITLCSSETNTDGGCVTGYDGACI 2340
Qy 1029 -----TGCTCTCCAAACT---ATGCCAAGC-----GG 1053
Db 2341 QSVFALNSSDPKVPKPIYSCADAFYTHSDCQIASSKCTTNGITGCTALGSCSYTAQAG 2400
Qy 1054 C-----GGCA-----GAGCTGGTCTTCTGCTCTCTCTGGAGAAGGTTC 1092
Db 2401 CYFNDKGAQYSSGAIVSTGVCTWDTISSCRDSCADLTGTTHATCSSQLSTCTSDGTTTC 2460
Qy 1093 TGTTCCTCTGAT---TTATGAAGCTC-----TA-----TAATAGATATAGG 1132
Db 2461 L-LKGVCTSYTTOICTACTTAVGSDGSCYWEIASATNNNTAKRLTCDIQNGTATNVCSV 2519
Qy 1133 TTTT---GTAC---CTT-TTTTAC-----AGGAGGTGACT--TTCT 1165
Db 2520 ALSTCVNSGTACIPKANGSTVTSKACNSGGLDGCIVFTQSTATGAAAGTCTCALMTACT 2579
Qy 1166 ---GTAACAA---TCCGATGTATATAACTTTTTATAAA 1199
Db 2580 IANNDQTAQAAARDCAWTAASGTGTAVASKCATHTCATNOA 2622

Db 827 GTGSGGAG--GSGGANFNNGTGTGTGTGTGGKG-----MGIIAGDGGPGGDG 874

Qy 1056 CGACAGCTGTGTCCTTCTGCTCCCTCGAGAAAGTTCTGTTCCTCCGTGAT--TTATGAAC 1113

Db 875 CNAGYGGKGG--TNGNGSGSGTGTGTGPGSGCAPTGTGTGKGAGGCGDGDGGAAT 932

Qy 1114 CTATAATAGACTATATAGGTTTTGTACCTTTTTTACAGGAAGGTGACTTTCTGTAACAAT 1173

Db 933 GVGDDGCGNG---GNGNGGTGTG-----SPGGLGGAGG-----TGLGGAGA 973

Qy 1174 GCGATG 1179 |||

Db 974 GGGADG 979 |||

RESULT 13

O76973 PRELIMINARY; PRT; 2721 AA.

AC O76973;

DT 01-NOV-1998 (TRENBLrel. 08, Created)

DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Surface antigen G.

GN 51G.

OS Paramesium tetraurelia.

OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;

OC Paramesium.

NCBI_TaxID=5888;

RN [1]_TaxID=5888;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=51;

RC MEDLINE=99038212; PubMed=9819394;

RA Duhaucourt S., Keller A.M., Meyer E.;

RT "Homology-dependent maternal inhibition of developmental excision of

RT Internal eliminated sequences in Paramesium tetraurelia.";

RL Mol. Cell. Biol. 18:7075-7085(1998).

DR EMBL; AJ010441; CAA09180.1; -

DR GO; GO:0005524; P:ATP binding; IEA.

DR GO; GO:0004672; P:protein kinase activity; IEA.

DR GO; GO:0007275; P:development; IEA.

DR GO; GO:0006488; P:protein amino acid phosphorylation; IEA.

DR InterPro; IPR003895; Paramesium SA.

DR InterPro; IPR003659; Plexin-like.

DR InterPro; IPR000719; Prot kinase.

DR InterPro; IPR001680; WD40.

DR Pfam; PF01508; Paramesium_SA; 34.

DR SMART; SM00639; PSA; 32.

DR SMART; SM00423; PSI; 2.

DR PROSITE; PS0107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS0676; WD_REPEATS_1; 1.

SQ SEQUENCE 2721 AA; 280343 MW; 630780EDB7876242 CRC64;

Query Match 11.9%; Score 894.5; DB 5; Length 2721;

Best Local Similarity 21.6%; Pred. No. 1.3e-44;

Matches 528; Conservative 51; Mismatches 607; Indels 1257; Gaps 127;

Qy 3 TCTGGG---GTGC-----TCCAGGAAAGCA--AATCTGGAGTT 40

Db 191 TKTGGGCVDSNNCSDTGLEIQCWNKLTATYWDGAACKDRICDNPATSLITDDACKT 250

Qy 41 AATGTTTGTAGTCATTTTAAAT-----CCTTG----- 69

Db 251 FRTDGTCTTKANGGCVTRTTCAAATAIOASCIKNSGGDCYWTGTACVDKTCANAPVTMT 310

Qy 70 ---CTG---GGGAGAGGCC-CCC-----TC---TCC----- 92

Db 311 NSACAGFVTGCTIKSGGGCVANGACSVANVOACVKNSNFDCIWDITCKEKT CANAPTI 370

Qy 93 -----CGGTAIC-----AGCGTTCTCATCTTTGAA-----TCGC-----GGC--- 129

Db 371 NNTHDLCSYLSLCTCVKSGGGCQNRTCANAPTLTTNDACSAYLTGNNTKTKGGCVTN 430

Qy 218 -----CGA--GGCGG-----GC-----CAATTG----- 234
Db 605 APKTTATTSDGCTYKSGCVANNPVNGSIQGGQDLPTTCAARKSTENCEBIETGPPPTCLWN 664
Qy 235 AATAAGAGGCGTGCTT-----CCAGCAGGCTCTATAAGTGAACCCCGCGG 282
Db 665 SATSACVKSCTASVTTTTGTPLTVFSPNTNCLAYLSNSACIANNTADGCIKPKSSCGSLT 724
Qy 283 CGAGGGTGC-----GGCGGTGAGTGC-----ACTGTAGCGAGTCTTTTGGTCTTTCTTTC 335
Db 725 TSNAGVSGSKANGDCVWNGTTCVDRTCNSNISLSTHSGCNGVINTC--TVNNAKTICOSLA 782
Qy 336 TCTTTGGGCGAC-----CTCTGAC--TC-----ATC-----CCCA 365
Db 783 TACTSYGSSENCKQTSAGKTCIWTGTACRNATCLDASDTNPSDGECSAYPTPSDTCV 842
Qy 366 GCATGAAGCGGTGAGC-----CCGGTGGCGGCTGCTAG-----AGCG----- 406
Db 843 LYKTGAV--GVVRSANCSDYVSOACVRLTAASDDCTWKSXSKCFSTNVLGACSTFQ 901
Qy 407 GTGTGC-----TGC-----CTGTGC--GAA--CGCAG--TCT-----GGCCAT 438
Db 902 GTKTLCEALKTGCTNAVGAVETDCTFCFVAVTGTSLTHAGCGGYSTTCTANSGGTACLT 961
Qy 439 CGCCC-----GGGC-----CGAGGAAGGC-----CGCGAGCTGAGAGCCGCT 480
Db 962 LAACSTYTAQGDVVRTDGLSCFWNAGSPAACQDITSTNCNALASLTGATHASCOAYS 1021
Qy 481 GAGCTTGCTGAGC-----GACATGAACAC-----TGCTACTCCG--CTGCGG 523
Db 1022 TA--CTSISDGTACQAIQAACSSYSALTACTKRTDGLKCFWNTGATPAACQDITSTNCA 1080
Qy 524 GAACTGTACCCGAGTCCGAGAG--GCACTAGCTT-----AGC-----CAGGTGA 570
Db 1081 LASLTGATHATCOAYSTACTSISDGTACQAIQAACSSYSALTACTORTDGLKCFWNTGAT 1140
Qy 571 AATCTACAGCGG-----TCATCGACTACATCT-----CGACTCGAGGTAGTCT 618
Db 1141 PAAACQDITSANCNALASLTGATHASCOAYSTACTSISDGTACQALQAACSSYSATCT 1200
Qy 619 GCGCGAGC-----CAGCCCT-----GGAC-- 638
Db 1201 QRTDGLKCFWNTAATPAACQDITSANCNLITGLSSATHATCOAYSTACVLAIGNAQOA 1260
Qy 639 ---CCCCTGATGCCCCCACCCTTC-----CAT--CCAG-----ACAGCC----- 673
Db 1261 FSTCEALTGSNLWTICQAFSTTCSVKRDGTGCVT:QSACTGYTTIANCYRSTAGYCTAN 1320
Qy 674 -----GAGCTC-----GCTCCGAACTGTGCTCT----- 698
Db 1321 SGDSACQAIASATTCEQIKLGSPAFDDTKNTFKTGICIALGTSGGCTKCAKNTTFFAH 1380
Qy 699 ---CCAACGACA--AAAGGAGCTTTTGCCACTGACTC-----GGCGGTGCTCT----- 741
Db 1381 SDCNTWLSCTSNVAVSPTACTTWAATCASLTTSTCVYAVEGCVVSGTSCVAKTCDTAS 1440
Qy 742 -----GACAC--CTCCGAAAGCGAGTGC-----TGG----- 766
Db 1441 ADTSFDSHAEICAYLSTCTVARTGCGQARATCASYSQCKFNSTGKCFWNNKTCV 1500
Qy 767 ---CGCCCGTT-----CTGCCTGGACCCCGGAACTCT-----CCTGC 803
Db 1501 DLNCGNIEATYTDHAECAVDVQLLCTVATNGAVPGCMARGACSSYSIEDQCTNP 1560
Qy 804 CGGAA-----GCGGAGCGGAGGATG 825
Db 1561 SGGVGVWNTNLTPVCODKSCCTAPTATPATHADCDYSFSTATIKCTVATPDNGGAVAL 1620
Qy 826 GGCCCCCACTT-----CG-----CCTGC-----CCACTTG 851
Db 1621 GCGCQTAACSSYTHQEQCRFNATGDLGNGTQCADKSCATAPATTDYDDNDKCRAFNN 1680

Qy 852 ACT-----TC-----ACCAATCCCT--TCCTGGAGACTAAAC 882
Db 1681 KCTVASSGGCGVDIPPTCBMTQKQCVSDTKGRSCVWNGTACITRTCNAPDSTASAEDC 1740
Qy 883 CTGGTCT-----C-----AGGAGGAGGAGCTGTGAACCTT--GT-----AGCC 919
Db 1741 NTYLAGCTLDSVKCKTKVCEDFAFATDALCKSALSTCTTNGTNCVTRGTCFQIQONAGCV 1800
Qy 920 TGAAGAGC-----CA-----GAGTAGCTC 939
Db 1801 TSSANVQCEWPAVGTNOAYCTVKTCTNATPATLTSASASYFTNCTTKNGGCVTKSTC 1860
Qy 940 T-----GGCC-----ACC-----AGCTGGCGAGC-- 959
Db 1861 SAVTIDVACTTALNGTVCAWDSAQNKCRDKODFSGTTTHAACQOTORAGCTAGANKCAR 1920
Qy 960 --TC-----ACCTGCTCCC-----ACCCACCCCA--AGTCTTAAGGT 995
Db 1921 VQNCBQTTIRACIEGTNGFCLNNVNSDGSKGACFRVTSCKSLANSDTSCKWISNQ 1980
Qy 996 CTTTTCAGAGC--GTGGAGGTGTGGAAGAGTGC-----TGCT----- 1032
Db 1981 CTTN--GSCIGITLTCSETNDDGCVSGYDGAACIOSVPALNSDPKVKPKYTSADAFY 2037
Qy 1033 ---CTCCAACT--ATGCCAAGCGGC-----GGC---AGAGCTGGTCTTCTGTCTC 1077
Db 2038 TTHSDCOTASKKCTTNGTTCGIALGACSSYTTQAGCYFNDKGALLTSGAITST--GICTW 2095
Qy 1078 CTTGAGAGAAAGTTCTTGTCCCTG--ATTATGAAC--TCTA-----TAATAGAGT-- 1125
Db 2096 DTTASSCRDQCADLTCTTHATCSSLSTCTSDGTSCLVKGACTSYTTTACTTAVGSDG 2155
Qy 1126 ---ATAT-----AGTTT-----TGTA-----CTT--TTT 1146
Db 2156 ICYWEALSTNNNTAKRLLACADIQNGTSTNVCAVALSCVNGTACIAKANGSTVTK 2215
Qy 1147 TAC-----AGGAAGTG-----ACTTCTGTACAA-----TCGGATGT 1180
Db 2216 TACNSGGLDGLICVFTQSTATGAAGTGTALMTACTTANSDOVACQAAKRCSTAAAGT 2275
Qy 1181 ATATTAACCTTTTATAAA 1199
Db 2276 GTTAVASKCATHTCATNOA 2294

RESULT 12
Q7TWB8
ID Q7TWB8 PRELIMINARY; PRT; 992 AA.
AC Q7TWB8;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE PE-PGRS family protein.
DE PE_PGRS57 OR NB3543.
GN Mycobacterium bovis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis".
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD95729.1; -.
KW Complete proteome.
SQ SEQUENCE 992 AA; 77324 MW; D0140B51C0042121 CRC64;
12.3%; Score 929.5; DB 16; Length 992;

Query Match

Qy	315	-GACTTCATT-----TTGGTTTTCTTTCTCTTTGGGCACTCTGG-----	353
Db	923	SGSSAAAAAARRAGHRAAGSAAAAAASGAGG-----SGGGYWGND	976
Qy	354	-----ACTCACTCCCACTGAAG-----CGCTGACCC	384
Db	977	GGYSDSAAAAAAGSGAGGAGDYGWGDGYSASAAAAAASGAGGAG	1036
Qy	385	GGTGCGCG-----CTGCTACGAGCGGTGTGCTGCTGTCCGACCGAGTC-----	431
Db	1037	GGYWGNDGGYSDSAAAAAAGSGAGGAGDYGSGSSAAAAAARRAGH	1096
Qy	432	-----TGCCCATCGCCGGCCGAGGAAGCCCGCAGCTGAGGAGCCGCTTGCTTGCT	489
Db	1097	DRAAGSAAAAAASGAGSGGYGWDGYSASAAAAAASGAGGAGV	1156
Qy	490	GGACG-----ACATGACACTGC-----TACTCCCGCTCGCGG	524
Db	1157	GGYWGNDGGYSDSAAAAAAGAGGAGGYGYGSDSAAAAAASG	1216
Qy	525	AACTGG-----TACCCGAGTCCCGAGG-----CACTCAGCTTA	580
Db	1217	AGGAGGYGWDGGYSDSAAAAAAGSGAGGAGDGGYSGSAAAAA	1276
Qy	561	GCCAGGTG-----AAATCTACCGGTC-----ATCGACTACA	596
Db	1277	ASGAGSGGYGWDGGYSDSAAAAAAGAGGAGGYGGYWGNDGGYSDSAA	1336
Qy	597	TTCTGACTGCAGGTAGTCTCTGCCGAGCCACCCCTCGACCCCTGA-----T	646
Db	1337	AAAAAASGAG-----AGGYGYGSDSAAAAAAGSGAGGAGGYWGND	1390
Qy	647	GGCCCCACCTTCCCATCCAGACCGAGCTCGCTCGGAACCTTGTCATCTCCAACGAC	706
Db	1391	GGYSDSAAAAAASGAGRGDGGYSGSSAAAAAARRAGYDRAAGSA	1450
Qy	707	AAAGAGCTTTTGCCACTGCTCGCGGTGTCTTACACCTTCAGAACGACGAGTGTGG	766
Db	1451	AAAAA-----AAAASGAGSGGYGWDGGYSDSAAAAAASGAGGAGG	1504
Qy	767	CCCGCTTCTGCTGGGACCCGGGAACCTCTCTCCCGGAAGCCGAGCGCAGGATGG	826
Db	1505	YG-GYSDSAAAAAASGSA-----GGAGGYGWDGGYSDSAAAA	1551
Qy	827	GCCCAACTT-----CGCCTGCCCACTTGACTTCACCAATCCCTTCTGGAGACTAA	880
Db	1552	AAAAAASGAGRGDGGYSGSAAAAAARRAGYDRAAGSAAAAA	1611
Qy	881	ACCTGTGCTCAGGAGCGAAGGACTGTGAACCTTGAGCCTGAAGCCAGAGCTAGCTCT	940
Db	1612	ASGAGSG-----GGYWGNDGGYSDSAAAAAASGAGGAGGYGYGSDSAAAA	1667
Qy	941	GGCCACAGCTGGCG-----ACGTCACC-----TGCTC	970
Db	1668	AAAAAAGSGAGGAGGYGWDGGYSDSAAAAAAGSGAGRGDGGYSGSSA	1727
Qy	971	CCACCCACCCCAAG-----TTCTAAGGTCTTTTCAGAGCTGG-----	1010
Db	1728	AAAAAARRAGHDRAAGCAAAAAAASGAG-GSGGYGWDGGYSDSAA	1786
Qy	1011	-----AGGTGGAAGAGTG-GCTGTCTTCAAACTATGCCAGCGCGGACAG	1061
Db	1787	AAAAAAGSGAGGAGGYGWDGGYSDSAAAAA-----AASGAGTGGYWG	1842
Qy	1062	CTGGTCTTCTGCTCTCTTGAGAAAGGTTCTGTGCCCTGATTTATGACTC-----	1114
Db	1843	DGGYSDSAAAAAASGAGGAGGYG-WBDGGYSDSAAAAAAGSGAGGR	1901
Qy	1115	-----TATATAGAGTATATAGTTTTGTACCTTTTTTACAGGAGTG-----	1158
Db	1902	GDGGYSGSAAAAAARRAGHDRAAGSAAAAAAGSGAGSGGYWGND	1961
Qy	1159	-----ACTTTCTGTAAATGCGATG-----TATATTAACTTTTAA	1198

Db	1962	GGYGSNSAAAAAAGSAGGAGGGYGGWGDGGYGSDSAAAAAAAAAAAAGSGAG	:
Qy	1199	AAG 1201	:
Db	2022	GAG 2024	:
 RESULT 11			
Q27167	PRELIMINARY;	PRT; 2395 AA.	
ID AC Q27167;			
DT 01-NOV-1996 (T-EMBLrel. 01, Created)			
DD 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)			
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)			
DB S1B type surface protein.			
OS Paramacium tetraurelia.			
OC Eukaryota, Alveolata, Ciliophora; Oligohymenophorea; Peniculida;			
CC Paramacium.			
OX NCBI_TaxID=5888;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=STOCK 51;			
RA MEDLINE=95098630; PubMed=7800503;			
RE Scott J., Leack C., Forney J.,			
RT "Analysis of the micronuclear B type surface protein gene in			
RL Paramacium tetraurelia.",			
RM Nucleic Acids Res. 22:5079-5084 (1994).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=STOCK 51;			
RA Scott J.M., Leack C.L., Forney J.D.;			
RL Genetics 133:189-198 (1993).			
DR EMBL; U07603; AAA81947.1;			
DR ENBL; L04795; AAA16710.1;			
DR PIR; S50820; S50820.			
DR GO; GO:0007275; P:development; IEA.			
DR InterPro; IPR002895; Paramacium SA.			
DR InterPro; IPR003659; plexin-like			
DR InterPro; IPR001680; WD40.			
DR Pfam; PF01508; Paramacium_SA; 31.			
DR SMART; SM00639; PSA; 28.			
DR SMART; SM00423; PSI; 3.			
DR PROSITE; PS00678; WD_REPEATS_1;			
SQ SEQUENCE 2395 AA; 246430 MW; 8632E02649B2D61 CRC64;			
 Query Match 12.3%; Score 930; DB 5; Length 2395; Best Local Similarity 23.4%; Pred.No. 9.8e-47; Matches 481; Conservative 67; Mismatches 637; Indels 874; Gaps 111;			
Qy 3 TCTGGGTG-CTGCCAGGAAGAACAAT-----TCTGGAAGTTAATGGTTTTGACT 53			
Db 248 TCASFRTDGTCTTKANGGVTRTCARAIQAACVNSSNGCYWTGTACVKTCTNA-- 305			
Qy 54 GATTTTAAATCCTTGCT---GGCGGAGA-----GGCCGGCTC---TC 91			
Db 306 PAIWTINSACAGFVTGCTIKSGGCVSNGACSAAVQAACVKNTSGTDICWDPTCKEKC 365			
Qy 92 C-----CCGGTATC-----AGCGTTCCTCATCTTTGA-----ATCC---GC 126			
Db 366 NAFTNNHDLCTSYLSICTVKSGGCGCPRICANAPTITTNDAEAFLPANNKITKSG 425			
Qy 127 GCC----TCC-----GCGGTCTTCGGGCTCAGACCAGCCGGAGAACCTGTTT 171			
Db 426 GGCVPNTCAVITLEAACVNSSGATCFWDSASSTCKDKTCAN-APSTNTNDLCVAFLS 484			
Qy 172 GCATTATAAGC-----GGGC-----TGTGAAC-- 193			
Db 485 SCTVNSTNAGCVDKTENSIAQTICDLDLNKACIKWKCYKECVLASSTATHADCO 544			
Qy 194 ---GCCCA---GGGC-----CGG-----CG-GGGCGGGGC--- 217			
Db 545 YDVGCTLSNTGTCVPLPKCEAITTEAACNRILQVTSVGKGYPLCGMNGSSCIDKAST 604			

QY	146	TCAGACGACCGGAGGAGCCTCTTTTCATTTTAAAG-----GGCGTGTGAACGCC	197
Db	524	AAAAAAAGSAGGAGGCGGCGGSSAAAAAASARRAAGHDSAAASAAAAA	583
QY	198	AGGCGCGCGCGGGCGGGC-----CGAGGCGGG-----C	227
Db	584	AAAG-AGSGGGYGMDCGGYSDAAAAAASGAGGGYGMDCGGYSDSA	642
QY	228	CATTTTGAATAAGAGGGTCCCTTCAGGCAGCTCTATAAGTACCGCGCGCGGAGC	287
Db	643	AAAAAASGARGSGG-----YGGYSDSAAAAAASAAAAAG-SGAGVGGGY	694
QY	288	GTGCGCGGTTCAGGTCACGTGATGAG-----CACTTCT-----	321
Db	695	GMGDC-GYSDSAAAAAASGARGSGGCGYGSAAAAAASAAAAAARRAGHD	753
QY	322	TTTGTGTTTCTTTCTTTGGGACCTCTGGACTCACTCCACAGCATGAAGCGCTGAG	381
Db	754	SAAGSAAAAAASGAGSGGGYGMDCGGYSDSAAAAAASAAAAAGSAGGAG	813
QY	382	CCGCTGCGCG-----CTGCTAGAGCGGTGTGCTGCTGTCTGGAACCGAG	429
Db	814	---GGYGMDCGGYSDSAAAAAASGARGSGYGGYSDSAAAAAASAAAAAG	870
QY	430	TCTGGCCATCGCCGGGCGGAGGCGCCGCGAGCTGAGAGCGGCTGAGCTTCT	489
Db	871	SGAGVGGGYGMDCGGYSDSAAAAAASGARGSGGCGY-----S	917
QY	490	GAACGACATGAACCA-----CTGCTACTCCGCTCGGGAACTGGTACCGAGTC	541
Db	918	GSAAAAAASARRAGHDSAAASAAAAAASGARGSGYGMDCGGYSD	977
QY	542	CCGAGAGGCACTGAGCTTAGCCAG---GTGG-----AAATCCTACAGC-CCGCTAT	588
Db	978	SAAAAAAASGAGGAGGGYGMDCGGYSDSAAAAAASGARGSGYGG	1037
QY	589	CGACTACATCTCGACCTGC-AGTACTCTTG---GCCGAGCGACCCCTGGACCCCT	643
Db	1038	YGSDSAAAAAASGAGGCGYGGYGMDCGGYSDSAAAAAASAAAAAG-----S	1091
QY	644	GATG-----GCCCCACCTTC-----CCATCCAGACCGCGAGCTCGC	681
Db	1092	GAGRGDGGYSGSAAAAAASARRAGHDRAAGSAAAAAASGAGSGGG	1151
QY	682	TCGGAACTGTTCATCTCCACGACAAAGGAG-----CTTTTGCCACTGAC	728
Db	1152	YGMDCGGYSDSAAAAAASGAGSGGGYGGYSDSAAAAAASGAGSGAG	1211
QY	729	TGCGCGTCTCTG---ACACTCCAGACGAGTGTCTGCGCCCGTCTGCTCGGAC	785
Db	1212	GAGGGYGMDCGGYSDSAAAAAASGAGSGGGYGGY-GYGSDSAAAAA	1270
QY	786	CCCGGAACCTCTCTGCGCG-----AAGCGGACCGAGGGATGG-GCC	829
Db	1271	AAAGSAG-----GAGGGYGMDCGGYSDSAAAAAASGAGRGDGGYGG	1322
QY	830	CCAACTTGCCCTGCCACTTGACTTCAACAATCCCTTCTGAG-----	875
Db	1323	SSAAAAAASARRAGHDRAAGSAAAAAASGAGSGGGYGMDCGGYSDS	1382
QY	876	-----ACTAAACCTTGCTG-----CTCAGGACGGAAG-----	902
Db	1383	AAAAAASGAGSGGGYGGYSDSAAAAAASGAGSGGGYGMDCGG	1442
QY	903	-ACTGTGACTGTAGCTGAAGACCGAGAGCTAGCTCTGGCCACCGCTGGCGAGTC	961
Db	1443	YGSDSAAAAAASGAGSGGGYGGYSDSAAAAAASGAGSGAGGAGG--	1500
QY	962	ACCTGCTCCACCCACCCCAAGTTCTTAAG-----TCCTTTTCA	1004
Db	1501	-----GYSDSAAAAAASGAGGAGGGYGMDCGGYSDSAAAAAASG	1555

Qy 1005 GCCTGCAGGTTCTGCA-----AGGAGTC 1027

Db 1556 GAGRGDGGYSGSSSAAAAAARRAGHPRAGSAAAAAASAGGAGGG 1615

Qy 1027 ---GCTGCTCTCAAACTATGCCAAG-GCGCGCGCAGAGCTGCTTCTTGCTCTCTGCTTGG 1082

Db 1616 YGWDGCGYSDSAAAAAASAGSAGGAGGCGWGDGYSAAAAAASAG 1675

Qy 1083 AGAAA-GGTTCTCTGCCCTGATTTATGACTC-----TATATAGAGT 1125

Db 1676 SGAGRGGGYWGDDGGYSDSAAAAAASAGSAGGAGGCGGYSAAAAA 1735

Qy 1126 ATATAGTTTGTACCTTTTTTACAGGAAG-----GTGACTTTCTGTACAAT 1173

Db 1736 AARRAGHRAAGSAAAAAASAGSAGGCGYSGWGDGYSAAAAAASAG 1795

Qy 1174 GCGAT-----GTATATATAAAGTATTTATATAAGT 1202

Db 1796 GAGSGGCGYGGYGGYSDSAAAAAASAGSAGS 1833

RESULT 10

Q964P4 PRELIMINARY; PRT; 2655 AA.

AC Q964P4

DT 01-DEC-2001 (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE Fibroin

OS *Antheraea yamanai* (Japanese oak silkworm).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;

OC Saturniidae; Saturniinae; Saturniini; Antheraea.

OX NCBI_TaxId=7121;

RN [1]

RP SEQUENCE FROM N.A.

RC TRANSPOSON-mariner-like element;

PA Lee J.-S., Hwang J.-S., Kang S.-W., Suh D.-S., Jin B.-R., Kim Y.-S.,

PA Lee K.-S., Goo T.-W., Yun E.-Y., Kim K.-Y.;

RT "Complete nucleotide sequence of fibroin gene cloned from *Antheraea yamanai*."

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF325500; AAK83145.1; -

SQ SEQUENCE 2655 AA; 218672 MW; 89BB810EF9CC4CA6 CRC64;

Query Match 12.9%; Score 970.5; DB 5; Length 2655;

Best Local Similarity 25.0%; Pred. No. 4,6e-49;

Matches 375; Conservative 54; Mismatches 721; Indels 353; Gaps 39;

Qy 1 GATCTGGG---GTGCTGCAGGAAAAAGCAATCTCG-----AGTTAATGG 45

Db 573 GAGSGGGYWGDDGGYSDSAAAAAASAGSAGGCGGCGGYSAAAAAASAG 632

Qy 46 TTTTG---AGTCATTTTAAATCTCTGTCGGCGAGCGCCGCTCTCCCGGTATCA 101

Db 633 AARRAGHRAAGSAAAAAASAGSAGGCGYWGDDG---YGSDSAAAAAASAG 690

Qy 102 GCGC-----TTCCTCATTTTGAATCCGGGGCTCCGGCTCTTCGGC 144

Db 691 GSAGGAGGYYWGDDGGYSDSAAAAAASAGSAG---GAGGYYWGDDGGYSDSAA 747

Qy 145 GTCAGACCAAGCCGAGGAGACCTGTTTGTCAATTAAGCGGGCTGTGACGCCAGCGCGC 204

Db 748 AAAAAAASAGSAGGCGGDDGGYSGSSAAAAAASARRAGHRAAGSAAAAAASAG 807

Qy 205 GCGGGCGGGCCGAGGCGGGCCATTTTCATTAAGAGCGCTGCCTCCAGCAGCGCTC 264

Db 908 AASGAGSGGGYWGDDGGYSDSAAAAAASAGSAGGAG---GDYWGDDGGYSG 863

Qy 265 TATAAGTGACCGCGCGGAGCGTGGCGCGTTGCAGGTTCACTGTAGCG----- 314

Db 864 DSAFAAAAAAASAGSGGGYWGDDG-YGSDSAAAAAASAGSAGGCGGCGY 922

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT	"Deciphering the biology of <i>Mycobacterium tuberculosis</i> from the
RT	complete genome sequence."
RT	Nature 393:537-544 (1998).
DR	ENSEL; AL022022; CAA17749.1; --
DR	PIR; B70807; B70807.
DR	Tuberculosis; Rv3512; --
DR	GO; GO:0004420; P:Hydroxymethylglutaryl-CoA reductase (NADPH). . .; IEA.
DR	GO; GO:0009058; P:biocytin synthesis; IEA.
DR	InterPro; IPR002202; HMG-CoA red.
DR	PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
DR	Complete proteome.
DR	NON_TER 1
SQ	SEQUENCE 1079 AA; 81163 MW; A79718DCB74B97D CRD64;

Query Match 12.9%; Score 971.5; DB 16; Length 1079;
Best Local Similarity 32.4%; Pred. No. 1.9e-49;
Matches 413; Conservative 19; Mismatches 522; Indels 321; Gaps 53;

6 GGGGTG---TGCCAG--GAAAAACAAATCTGGAGT-----TAATGTTTGGAGT 53
Qy
18 GNGGNGADTTTAAAGTTTGGAGGAGGAGCTGCTGGAGTGTGGGNGGNGGNGGTG-GK 76
Db
54 GATTTTTAAATCCTTGTGGGGCGAGAGCCCG-CCTCTCCCGATCAGCCCTTCCTCA 112
Qy
77 GTTGCGDGA-----GSSGGAGGKGGNGGAGKAGTGSAPGTAGTGCGGGGNGGIGAA 131
Db

113 TTCTTTGAATCCGGGCTCCGGGCTCTTGGGCTCAGACACCGCGAGAGACCCTGTTTG 172

Qy

132 GTTGPVGTCAAGGTCGGG--GAGGT---GGDGGANNGTAG--AGAGGNGG--KGGDGG 182

Db

173 CAATTTAACGGGGCTGTGAACGCCCAAGG-----CCGGCGGGGGCGGGCCGAGGCGGGCC 228

Qy

183 AGVTSSTAGNSGGAGSGSGKGGDAGAGGAGATTPANGIANGNGDGGDGAAGAVGTSG--- 239

Db

Qy	229	ATTTTGATAAAGAGCGCTGCTTCCAGGCAGGCTCTATAGTGTACCGCCGCGCGCA--G	288
Db	240	-----ATGAGDGGHGTGAAGNGGTGGAGSGIDGVGGTG--GTGNGNGNGALGG	289
Qy	287	CTGCGCGCGCTTCAGGTCACGTCTAGCGGACTCTTTTGGTTCTTTCTCTTTTGGGCA	346
Db	290	AGDAG- : : : : : -GSGNSGNGGTTGGKGNAGAGCA-----AGNNGTV-----GNGTGGD- -	335

	QY	347	CCTCTGGA	TACTCC	CAGCAT	AAGA	GCGTGTAG	CCCCGGTG	----	CGCGGTGTC	TAC-	400
			:	:	:	:	:	:	:	:	:	
	Db	336	---	GNNGA	AAATA	AGSNG	GAGTGS	AGGNGTGTGR	GGSGAG	DGDTGC	GVGGKGG	391
			:	:	:	:	:	:	:	:	:	
	QY	401	---	GAGG	CGGT	-----	GTCTC	CTCTGCG	GAAC	CAATC	-----	431
			:	:	:	:	:	:	:	:	:	
	Db	392	EVGAG	AGAG	SGSPNTS	PCGNG	CGQG	GGSGAG	AGAG	-AGAGG	HANFTATN	450

488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

[illegible][illegible]

775	Qy	CTGCTCGGAGCCCGGAACTCTCTCGCGGAGCCGGA	CGGCAGGAGTGGGCCCCAAC	834
760	Db	GTGSGGTGSDGNGNGADNTAN-MTAQAGDGG--	NGDGGFGGAGAGGG-----	809
835	Qy	TTGCGCCCTGCCACTTCACTTACCACAAATCCCTT	CTCGAGACTAAACCT-----	884
810	Db	-----GLTAGANGTGGGAGG-----	DGNGNAIGHGGLTDDPGNGGTG	850
885	Qy	--GGTGCTCAGAGCGGAAGCACTGTGA----	ACTTGTAGCTGTGAAGCCAGACTAGCT	938
851	Db	CNGGTGGT--GGAGTGLGGTGGDNGNGNGGTGG	EGEYVGGAGGTGGAAGNGDGG--	906
939	Qy	CTGGCCACCAGCTGGGGGACGTCA	CCCTGCTCCACCCACCCCAAGTCTTAAGGTC	998
907	Db	TGG-----TGGDG-----	GAGGTGGT	923
999	Qy	TTACAG-----CGTGGAGGT-GTGGAAGAGT	GGCTGCTCTCAAACTATGCCAAGCGG	1053
924	Db	GGTGLGDPYVGGSGDGTGGSGGAAGNGGNGAG--	-----AGNGNGG	968
1054	Qy	CGCGAGACTGGTCTTCTGGTCTCTCTTGGAGAAG	GGTCTGTGCGCTGATTTATGA	1113
969	Db	TTGAGGIGGTG-----NGGDABEPGVP	PAGGAGGAGTTGGKGG---TGGNGSGTG---	1016
1114	Qy	CTATAATAGATATATAGGTTTTGTATACCTTTT	TTTACAGGA-----AGGTGACTTTCTG	1166
1017	Db	-----SGTGDGCTGGGGNGGTGNGKGD	TSGGGAGDGGKAPAGGTG-----	1062
1167	Qy	TAACAATCGGATGTA	1181	
1063	Db	-----GAGDGGGA	1070	

RESULT 9

076786
ID 076786 PRELIMINARY; PRT; 2639 AA.

AC	O76786;
AD	01-NOV-1998 (TREMBlrel. 08, Created;
AE	01-NOV-1998 (TREMBlrel. 08, Last sequence update)
AF	01-NOV-1998 (TREMBlrel. 08, Last sequence update)
AG	01-OCT-2003 (TREMBlrel. 25, Last annotation update)
AT	Fibroin.
DE	Anthraea pernyi (Chinese oak silk moth).
EE	Anthraea pernyi (Chinese oak silk moth).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC	Sturniinae; Saturniinae; Saturniini; Anthraea.
OC	NCBI TaxID:7118.

OK
RN
[1]
MCEI TAXID~ / 1115;

RP SEQUENCE FROM N.A.

RX MEDLINE=20496953; PubMed=11040284;

RA Sezutsu H., Yukuhiro K.;

Dynamic rearrangement within the

RI. J Mol Evol 51:329-338 (2000)

DR EMBL; AF083334; AAC32606.1; -;

DR PIR; T31328; T31328.

SQ SEQUENCE 2639 AA; 216057 MW; 2

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Query Match	Best Local Similarity	Score
	25.18	12.98

DEB: LOCAL SIMILARITY 23.1%, FIDUCIAL MATCHES 386: Conservative 60: Mi

.....

Qy 2 ATCTGGGTGCTGCCAGGAAAAGC

7

Db 344 AASSGAGGRGDDGGYGS GGSSAAAAA

52 -----GTGATTTTAAATC

2
3

D**b**
404 GSGGYGWGDGGYGSDSA^{AAAAA}

1000

87 CTCCTCCCGGTATCAGCGCTTCCTC

464 A A A A S G A G G S G G Y G G Y G S D S A A A A

[illegible]

QY 1147 TACAGGAGGTGACTTTCTGTAAACAATCGGATGTATAT-----TAAACTTTTAA 1195
 Db 1584 GSGDGGAGNGGAACTGCTGGDGLTGTGTGGGGTGGDGGNGGNGANDTANMTA 1642
 QY 1196 TAAAG 1201
 Db 1643 QAGSDG 1648

RESULT 6

Q7TWC3 PRELIMINARY; PRT: 1460 AA.
 AC Q7TWC3;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PE-PERS family protein.
 GN PE-PERS4 OR MB3538.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RA MEDLINE=22709107; PubMed=12789972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrall B.S., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248346; CAD95724.1; -.
 KW Complete proteome.
 SQ SEQUENCE 1460 AA; 113187 MW; 1806DB7DD4BDBA9 CRC64;

Query Match 13.2%; Score 998.5; DB 16; Length 1460;
 Best Local Similarity 31.0%; Pred. No. 6.5e-51;
 Matches 429; Conservative 9; Mismatches 576; Indels 371; Gaps 52;

QY 6 GGGGTGCTGCCAGGAAAAGCAAACTCTGAAGTTAATGTGTTTG-----AGTCATTT 58
 Db 161 GNGGFGAGGGGGGTGGAGGAGGWLFGVGGAGGVGGAGGTGGAGGPGGLNGGGAGGV 220
 QY 59 TTAATCTTGTCTGCGG-----AGAGCCCGCTCTCCCGGTATCAGCGCTTCT 110
 Db 221 GGAG-----GGTGAGGAAELFFGAGGAG-----AGGAGTDGSPGATG--- 259
 QY 111 CATCTTCTTAATCCCGGCTCCGGGTCTTCGGGTTCAGACCCGAGGAGGAGCCTGT- 169
 Db 260 -----GTGGHGVGGDGGWLPFGAGGAGGAGGAGGAGSDGGLGGTGTG 305
 QY 170 -----TTGCATTTAAGCGGCTGTGAAGCCGAGG-GGCGCGGGGCGG-GGCCGA 220
 Db 306 GTGGAGGAGGAGLLGGGGGGLGAGAGGCGGTGGAGGDLVGGVGGTGGKGVGVAGL 365
 QY 221 GCGCGGCCATTTTGAATAA-----AGAGGGGTGCTCTCCAGGAGGAGCTCTAAGT 271
 Db 366 GGAGGAAGQLFSAGGAAGAVGGTGGGGAGGAGGAGGAGNAGSAGADGAGGTGNAGA 425
 QY 272 GACCGCCGCGGAGCGTGGCGGTTCAGGTCACTGTAGCGGACTTCTTTTGGTTT- 330
 Db 426 GGAGGAAGTGGTGGVGAAGKAGTGGTGGG-----AGGAGSAGTATATATGATGGTGG 481
 QY 331 -----CTTTCTCTTTGGGGCACCTCTGGACTCACTCCCGGAGCATGAAGGCGC---TG 379
 Db 482 GAGGAGGAGGNTGVGGTNGSGGGGTGGAGG-----GGAGGVGADNPTG 526
 QY 380 AGCCCGGTGGCGGTGCT--ACAGGCGG-TGTGCTGCTGTCTCGGAAGCGACTCTG--- 433
 Db 527 TG-GAGGTG-GAGG-TGGTGAAGAGGAGGAVGTGGTGGVGVVGVNAGTGGTGGKGGAG 583
 QY 434 -GCCATCGCCCGGGCCGAGG-----GAAGGCGCGGCGAGCTGAGGAGCGGCTGAGCT--T 486

RESULT 7

Q8VIZO PRELIMINARY; PRT: 1715 AA.
 ID Q8VIZO;
 AC Q8VIZO;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE PE-PGRS family protein.
 GN MT3612.1.

Db 584 TGFAGGAGGGGGSSGAGGTNGSGGAGG---TGQGGAGGAGGAGADNPITGIGAGGT 640
 QY 487 GCTGGACGACATGAACACACTGCTACTCCCGCTTCGCGGAAC--TGATACCCCGAGTCCCG 544
 Db 641 GGTGGAAGAGGAGGA--IGTGGT-----GGAVGSVGNAGIGGTGGT---GGVG---G 684
 QY 545 AGAGGCACCTCAGCTTAGC-----CAGGTGGAATCTCTACAGCG 582
 Db 685 AGGAGAAAAGSSATGAGFAGGAGGAGGNSGVGGTNGSGAGAGGKGGTGGAG-G 743
 QY 583 CGTCATCGACTACATCTCTGACCTCGAGGTAGTCTTGGCCGAGCCAGCCCCCTG----- 635
 Db 744 SGADNPTGAGFAGGAGGTGGAAGAGGAGGATGTGGTGGVVGATGSAIGAGAGRGDGDGD 803
 QY 636 ----GACCCCTGTATGGCCCCCACCTTCCATCCACAGACAGCGAGCTCGCTCCGAACT 690
 Db 804 GASGLGLSLGDFDGGG-----QGGDGSAGAGGAGGAGGDDGD 846
 QY 591 TGTCACTC-----CAACGACAAAG-----GAGCTTTTGCCACTGACTCGGCGGTGT 738
 Db 847 GGDGATGAAGLDNGVGGPAGGAGGAGNGNAGVGLTAKAGDGGAGNGNGAGGAGG 906
 QY 739 C-----CTGACACCTCCAGAACGACAGGTGCTGGCCGC---CGTCTGT 777
 Db 907 AGDNFNGGGAGGGGGGGLGGAATTSINAGGAGNGGTGGKGGAGGAGTGLGVGSSG 966
 QY 778 CTTGGACCCCGGAGACCTCTCTCGGAGAGCCG-----GACGGCAGGAGT----- 825
 Db 967 GTGDDGDDAGAGGG-----GGFGAAGKAGGAGGNGVGGDGGEGATGLGLDLGFG 1016
 QY 826 ----GGCCCCAACTTGCCCTGCCCTGACCTTCACTTCAACAAATCCCTTCTCGAGAGCTAAA 881
 Db 1017 AGGGGGGGAGGNAGAG-----GINGAGGTGG-----TGAGGDDGAPA 1054
 QY 882 CTTGCTGCTCAGGAGCGAGG-----ACTGTGAATCTTAGCTGAAGAGCCAGAGCTA 935
 Db 1055 TLIGPDPGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1111
 QY 936 GCTCTGCCACACAGCTGGGCGAGCTCACCTGTCTCCACCCACCCCAAGTCTTAAGT 995
 Db 1112 DPSIDGG-----QGGAGHGGGGKGLNSTG-----CTGGAAGGAGTGGC--TGCTCTCAAACT 1041
 QY 996 CTTTTCAGAGCGTGGAGGT-----GNGGAGGAGGNGGDDGFGSGGTGGTGGAGGAGGAGTGGT 1190
 Db 1147 -----GNGGAGGAGGNGGDDGFGSGGTGGTGGAGGAGGAGTGGT----- 1190
 QY 1042 ATGCCAGGCGCGGAGAGCTGGTC---TTCTGTCTCTCTTGGAGAAAGTCTCTGTTCG 1098
 Db 1191 -AGNAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1243
 QY 1099 CCTGATTTATGACTCT-----ATAATAGATATATAGGTTT 1136
 Db 1244 GTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1303
 QY 1137 GTACCTTTTTCAGGAAGGTGACTTTCTGTAACTATGATGTATATTAACACTTTTAT 1196
 Db 1304 G-----GTNFGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1358
 QY 1197 ABAAG 1201
 Db 1359 GGLGG 1363

QY 552 CTCAGCTTAGCCAGGTGGAAATCTCTACAGCGGTCATCGACTACATTTCTCGACTGCAAG 611
 Db 747 SGGAGGTGGAGGAGGAGADNPFTGIGTG-GGGTGGAAAGAGGAG---GAAGTGGTGG 802
 QY 612 TAGTCTGCTGCCAGGACGACCCCTGACCCCTGATGGCCCCCACCCTTCCCATCCAGACAG 671
 Db 803 MGTTCGAGGAGGAGGAGGAGGAGGAGADDPGATGG-----TGFAAG 845
 QY 672 CCGAGTCTCCTCCGGAATTTGTCATCTCAAGCAGAAAGAGAGATTTTTCATGACTCG 731
 Db 846 AGAGGGGSSCAGTNGSGGAGTGG-----QVAVAGAGISFNSNGSGTGG--TG 895
 QY 732 GCGGT-----GTCTGACACCTCCAGAACGCAAGTCTGCGCCCTCTCTGCTGGGACCC 787
 Db 896 GVGCTGGGAGTGA-----GDPKGTGTGTGTGGSG-----GAGSGGA--N 938
 QY 788 CCGGAACCTCTCTCCCGGAGCCG-----GACGCGAGGATGGGCCCCCAAC 834
 Db 939 FNGG-----TGTGTGTGKGLNDTGLSATGTTGTGTGGTGGTGG-----981
 QY 835 TTGCGCCCTGCCACTTGACTTCCACCAATCCCTTCTGAGACTTAAA--CCTGTGTCTCA 892
 Db 982 -----AGDASAGTGG-----TGGAGGNAGAGLANTGGT---A 1012
 QY 893 GAGCGGAG-----ACTGTGACTTTGAGCTCTGAAGCCAGAGCTAGCTCTGGCCA 945
 Db 1013 GNAGTGGGGGAGGGGSGGLGQPFAGAGAGGAGGAGGAGGAGTNG---SGG---1066
 QY 946 CCAGCTGGCGGACGTCACCTCTCTCCACCCCAACCCCAAGTCTT-AAAGTCTTTTTCAGA 1004
 Db 1067 --AGGAGGCGGAGGAGISFNSGNS-----GGTGTGGTGGTGGGAGT 1109
 QY 1005 GGTGTGAGGT-CTGGAGAGGTGCTGCTCTCCAACTATGCCAAGCGCGGCGAGAGCT 1063
 Db 1110 GAGDPKGTGTGTGTGGGAGGSGGANFNGTGGTGGTGG--GTGKGGMGGIAGDGP 1167
 QY 1064 GGTCTTCTGCTCTCTGGAGAAAGTCTCTGTGCTCTGATTTATGAATCTATAATAGA 1123
 Db 1168 GG-----DGGN--AGVGKGTGNGSGSGGTGTGGAGGAGGAGLANT-GGTAGNAGI 1218
 QY 1124 GTATATAGTTTGTACCTTTTACAGGAG-----GTGACTTCTGTACATGCGATG 1179
 Db 1219 G-GGGGGGNGGSGSGGLGQPFAGGPGKGGAGNAGTGTGNGSAGGAGGAGGAG 1277
 QY 1180 TATATTAAATTTTATATAAGT 1202
 Db 1278 GAGISFNSNGTGTGTGGTGT 1300

RESULT 4

Q8VIY9 PRELIMINARY; PRT; 1217 AA.
 AC Q8VIY9;
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
 DE PE PGRS family protein.
 GN MT3615.3.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1173;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB007164; AAK47974.1; --.
 DR TIGR; MT3615; --.
 DR InterPro; IPR002952; Eggshell.
 DR InterPro; IPR000084; PE region.
 DR InterPro; IPR002173; PFKB.
 DR Pfam; PF00934; PE; 1.
 DR PRINTS; PR01228; EGGSHLL.
 DR PRODOM; PD001223; PE region; 1.
 DR PROSITE; PS00583; PFKB_KINASES_1; 3.
 SQ SEQUENCE 1217 AA; 94463 MW; 2F5BDCA5A30A056E CRC64;

Query Match

Best Local Similarity 33.1%; Score 1035; DB 16; Length 1217;
 Matches 409; Conservative 18; Mismatches 512; Indels 298; Gaps 48;

QY 1 GATCTGGGCTCTGCCAGGAAAAGCAATCTTGA--AGTTAATGTTTG--ACTGAT 56
 Db 208 GGLTWGGGAGGAGGAGGTGGAGRAELLFAGAGGAGGATGGGATGGTGGHGVGGD 267
 QY 57 TTTTAAATCTTCTGCGGAGAGGCCCTCTCCCGGTATCAGCGTCTCTCATTTCT 116
 Db 268 GGLAPGAGAGGAGGAGGAGGAGSDG-----GALGTTGGTG-----303
 QY 117 TTGAATCCGGGCTCCCGCTCTTCGCGCTCAGACCCAGCGGAGGAGCCTGTTTGCAT 176
 Db 304 TGGAGAGRGALLLGAGGQ-----GGLGAGGQ-----GMGAGGAGANPFIGT 352
 QY 177 TTAAGCGG--GCTGTGAACCCAGGCGCGCGGCGCGGCGCGGCGCGGCGGCGGCG 231
 Db 353 GGDGTTGSGSAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 412
 QY 232 TTGAATAAAGAGGCG--TGCTTCCAGGAGGCTCTATAGTGACCGCGCGCGGAGCCT 289
 Db 413 GTGFAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 471
 QY 290 GCGCGCTTTCAGGTCTCTGTAAGGAGTCTTTTGGTCTTTCTTTCTTTTGGGCACT 349
 Db 472 ----GIGTGTGGGTGGAAGAGGAGGA-----AGTGTGTMGTGTTNAGVGGAGG 522
 QY 350 CTGACTCACTCCCGCAGATGAGGCGCTGAGCCCGTG--CGCGCTCTCTACGAGCGG 407
 Db 523 GAGAGADA-----DQFATGTTGTFAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 572
 QY 408 TGTCTCTCTGTGGAAACGAGTCTGCGCATCCCGCGCGCGCGGAGGAGGAGGAGGAG 467
 Db 573 QGGG-----AGSSSAGGTTNGSAGGAGT--GCGGAGGAGGAGGAGGAGGAGGAG 620
 QY 468 CTGAGGAGCGCTGAGCTTCTGACGACATGAACCACTGCTACTCCCGCTGCGGAGAA- 526
 Db 621 DGTTGGAGAGGAGGAGGAGTGTGTGTMIG--TTGNAGVGGAGGQ-----GGD 572
 QY 527 -----CTGGT-----ACCGGAGTCCCGAGAGCG--ACTCAGCTTACCGAGTGGAAATCC 575
 Db 673 ADQFATGTTGTFAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 732
 QY 576 TACAGCGCTCATCGACTACATCTTCGACTCGAGTAGTCTCGGCGGAGGAGGAGGAGGAG 635
 Db 733 TGI GTG-GDGGTGGAGGAG-----GGAAGAGTGTGTGTMIGTTGNAG-----VG 775
 QY 636 GACCCCTGTATGGCCCCCACTTCCCATCCAGACACCGCA-----GCTCGCTCCGGAATC 690
 Db 776 GA-----GGGG-----DGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 808
 QY 691 TGTCTATCTCAACGACAAAAGAGAGCTTTTGCCTGCTGCGCGCTGTCTTCCACCTCC 750
 Db 809 -----GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 850
 QY 751 AGAAGCAGAGGTCTGGCGCCGCTTCTGCTCGGAGCCCGGAGGAGGAGGAGGAGGAGGAG 810
 Db 851 -----GNGTGTGGTGGV-----GTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 899
 QY 811 CG-----GACGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 866

[illegible]

PRT; 3145 AA.

RP SEQUENCE FROM N.A.

RA Kaneko T., Nakamura Y., Sato S., A

RA Kishida Y., Kiyokawa C., Kohara M.
 22 Mochizuki Y. Nakayama S. Nakazaki

RT "Complete genome structure of the

DNA Res. 7:331-338 (2000).

DR pfam; PF01391; Collagen; 38.

SQ SEQUENCE 3145 AA; 271616 MW; 271616 MW;

Best Local Similarity 36.1%; Pred.

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840. 84

858 GATGATGATGATG--ATGATG

57 TTTTAAATCCTTGC TGGCGGAGAG

22

970 ATCTTCATGSTGATGTT--GATGAC

QY 171 ----TGCAATTTAAGCGG-----GC

DD
1027 ATGD TGA TGS TGATGATGATGATG

[illegible]

269 A GTGACCGCCGCGGCGAGCGTGCGG

Db
1146 GDTGATGSTGATGATGATGTGTC

QY
318 TCTTTGGTTCCTCCTCC

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 15:03:25 ; Search time 57 Seconds
(without alignments)
6659.095 Million cell updates/sec

Title: X69111
Perfect score: 7544
Sequence: 1 GATCTGGGCTGCCAGCA.....TTAACTTTTATAAAGTT 1203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 107041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 11 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 1268.5 | 16.8 | 2147 | 16 Q98MGB | Q98MGB rhizobium 1 |
| 2 | 1230 | 16.3 | 3145 | 16 Q98MG7 | Q98MG7 rhizobium 1 |
| 3 | 1074.5 | 14.2 | 1489 | 16 O53559 | O53559 mycobacteri |
| 4 | 1035 | 13.7 | 1217 | 16 Q8VY9 | Q8VY9 mycobacteri |
| 5 | 1016 | 13.5 | 1938 | 16 Q7TWC0 | Q7TWC0 mycobacteri |
| 6 | 998.5 | 13.2 | 1460 | 16 Q7TWC3 | Q7TWC3 mycobacteri |
| 7 | 979 | 13.0 | 1715 | 16 Q8VIZ0 | Q8VIZ0 mycobacteri |
| 8 | 971.5 | 12.9 | 1079 | 16 O53557 | O53557 mycobacteri |
| 9 | 971.5 | 12.9 | 2639 | 5 O76786 | O76786 antheraea p |
| 10 | 970.5 | 12.9 | 2655 | 5 Q964F4 | Q964F4 antheraea y |
| 11 | 930 | 12.3 | 2395 | 5 Q27167 | Q27167 paramecium |
| 12 | 929.5 | 12.3 | 992 | 16 Q7TWC8 | Q7TWC8 mycobacteri |
| 13 | 894.5 | 11.9 | 2721 | 5 O76973 | O76973 paramecium |
| 14 | 889.5 | 11.5 | 1384 | 16 Q8VIZ1 | Q8VIZ1 mycobacteri |
| 15 | 867.5 | 11.5 | 2717 | 5 Q94710 | Q94710 paramecium |
| 16 | 864.5 | 11.5 | 1360 | 16 Q7TWC4 | Q7TWC4 mycobacteri |

```

17 863 11.4 988 5 O17434
18 860.5 11.4 1574 11 O88281
19 860 11.4 1381 16 O53552
20 852 11.3 1306 16 O53775
21 848 11.2 1306 16 Q7U1Q7
22 817 10.8 1507 16 Q8VJ23
23 807.5 10.7 1468 5 Q9GUB5
24 800.5 10.6 1665 16 O53215
25 797 10.6 1538 16 O53395
26 793.5 10.5 1408 16 O8VK17
27 789.5 10.5 773 16 Q7U160
28 789 10.5 1329 16 O68810
29 783.5 10.4 1246 4 O75095
30 779.5 10.3 1408 16 Q7U022
31 772 10.2 749 16 O53844
32 747 9.9 909 16 Q7U1D3
33 739.5 9.8 13288 6 O18758
34 726.5 9.6 1150 16 Q7TYG8
35 726 9.6 868 16 Q7U2D8
36 717 9.5 882 16 O53845
37 716 9.5 2189 5 Q9BI05
38 712 9.4 1664 5 Q9TV42
39 712 9.4 2360 5 Q7Y2P0
40 709.5 9.4 879 16 Q8VKD2
41 706.5 9.4 853 16 O53439
42 701 9.3 850 16 Q7U0P1
43 687.5 9.1 837 16 O53684
44 685 9.1 831 16 Q7U2D6
45 684.5 9.1 2352 5 O61240

```

ALIGNMENTS

RESULT 1

```

Q98MGB PRELIMINARY; PRT; 2147 AA.
ID Q98MGB
AC Q98MGB, 2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical glycine-rich protein mlr0585.
GN MLR0585.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; Pubmed=11214968;
RA Kaseko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002985; BAB48145.1;
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 27.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2147 AA; 181047 MW; FE2191AA84089AB CRC64;

```

```

Query Match 16.8%; Score 1268.5; DB 16; Length 2147;
Best Local Similarity 35.6%; Pred. No. 1.4e-66;
Matches 460; Conservative 64; Mismatches 628; Indels 141; Gaps 58;
QY 1 GATCTGG--GTCGTGCCAG-GAAAAACCAATCTG-----GAAGTTAATGCTTTTGA- 51
DB 702 GATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 761

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:51:29 ; Search time 18 seconds
(without alignments)
3480.022 Million cell updates/sec

Title: X69111

Perfect score: 7544
Sequence: 1 GATCTGGGGTGTGCCAGGA.....TTAAACTTTTATAAAGTT 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 1 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 970.5 | 12.9 | 1901 | 1 YZ08 MYCTU | O53553 mycobacteri |
| 2 | 962.5 | 12.8 | 5263 | 1 PBOH_BOMMO | P05790 bombyx mori |
| 3 | 891.5 | 11.8 | 2715 | 1 G156_PAPER | P13837 paramedium |
| 4 | 857 | 11.4 | 2704 | 1 G168_PAPER | P17053 paramedium |
| 5 | 784.5 | 10.4 | 914 | 1 W422 MYCTU | O06794 mycobacteri |
| 6 | 714.5 | 9.5 | 957 | 1 Y278 MYCTU | P56877 mycobacteri |
| 7 | 676.5 | 9.0 | 2531 | 1 NTC1 RAT | Q07008 rattus norv |
| 8 | 674.5 | 8.9 | 2556 | 1 NTC1 HUMAN | P46531 homo sapien |
| 9 | 670.5 | 8.9 | 5376 | 1 ZAN_MOUSE | O88799 mus musculu |
| 10 | 668 | 8.9 | 778 | 1 YQ34 MYCTU | P71933 mycobacteri |
| 11 | 666 | 8.8 | 481 | 1 LORI_MOUSE | P18165 mus musculu |
| 12 | 661.5 | 8.8 | 2531 | 1 NTC3_MOUSE | Q01705 mus musculu |
| 13 | 656 | 8.7 | 2321 | 1 NTC3_HUMAN | Q9um47 homo sapien |
| 14 | 654.5 | 8.7 | 1700 | 1 BAR3_CHITE | Q03376 chironomus |
| 15 | 640.5 | 8.5 | 2319 | 1 NTC3 RAT | Q9r172 rattus norv |
| 16 | 638.5 | 8.5 | 801 | 1 Y747 MYCTU | O53810 mycobacteri |
| 17 | 637.5 | 8.5 | 2318 | 1 NTC3_MOUSE | Q61982 mus musculu |
| 18 | 631 | 8.4 | 747 | 1 SPD1_NEPCL | P19837 nephila cla |
| 19 | 616.5 | 8.2 | 2437 | 1 NTC1_BRARE | P46530 brachydanio |
| 20 | 613 | 8.1 | 2524 | 1 NTC1_XENIA | F21783 xenopus lae |
| 21 | 603.5 | 8.0 | 672 | 1 PHX5_MOUSE | P08399 mus musculu |
| 22 | 588 | 7.8 | 2703 | 1 NOTC_DROME | P07207 drosophila |
| 23 | 583 | 7.7 | 2471 | 1 NTC2_HUMAN | Q04721 homo sapien |
| 24 | 579 | 7.7 | 1964 | 1 NTC4_MOUSE | P31695 mus musculu |
| 25 | 577.5 | 7.7 | 1046 | 1 PSTA_DICDI | P11976 dictyosteli |
| 26 | 577.5 | 7.7 | 2471 | 1 NTC2 RAT | Q9gw30 rattus norv |
| 27 | 574 | 7.6 | 2470 | 1 NTC2_MOUSE | Q5516 mus musculu |
| 28 | 557 | 7.4 | 4289 | 1 TENX_HUMAN | P22105 homo sapien |
| 29 | 540 | 7.2 | 2907 | 1 FEN2_MOUSE | Q61555 mus musculu |
| 30 | 535.5 | 7.1 | 2003 | 1 NTC4_HUMAN | Q99466 homo sapien |
| 31 | 535 | 7.1 | 1064 | 1 FEP1_STRPU | P10079 strongyloce |
| 32 | 527 | 7.0 | 2911 | 1 FEN2_HUMAN | P35556 homo sapien |
| 33 | 514.5 | 6.8 | 641 | 1 EBN1_EBV | P03211 epstein-bar |

34 475 6.3 603 1 YD25 MYCTU Q10637 mycobacteri
35 474.5 6.3 2871 1 FBNI_BOVIN P98133 bos taurus
36 469.5 6.2 2871 1 FBNI_HUMAN P25555 homo sapien
37 460 6.1 2871 1 FBNI_MOUSE P21483 mus musculu
38 450.5 6.0 349 1 GRP_ARATH P21483 arabidopsis
39 447 5.9 2871 1 FBNI_PIG Q95V36 sus scrofa
40 445.5 5.9 498 1 YI18 MYCTU Q50615 mycobacteri
41 438.5 5.8 384 1 GRPI_PETHY P09789 petunia hyb
42 435.5 5.8 860 1 ELS_MOUSE P54320 mus musculu
43 435 5.8 713 1 TSA4_GIALA P21849 giardia lam
44 433 5.7 194 1 KRUE_HUMAN O75690 homo sapien
45 433 5.7 465 1 GRP2_PRAVU P10496 phaseolus v

ALIGNMENTS

RESULT 1

YZ08 MYCTU STANDARD; PRT; 1901 AA.

AC Q53553; 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv3508 precursor.

GN Rv3508 OR MTW023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RI complete genome sequence."
RL Nature 393:537-544 (1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL; AL022022; CAA17745.1; ..

DR PIR; F70806; F70806.

DR Tuberculist; Rv3508; ..

DR InterPro; IPR000084; PE_region.

DR Pfam; PF00934; PE; 1.

DR ProDom; PD001223; PE_region; 1.

KW Hypothetical protein; Repeat; Signal; Complete proteome.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN

FT RV3508.

SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match 12.9%; Score 970.5; DB 1; Length 1901;

Best local similarity 31.9%; Pred. No. 1.3e-42;

Matches 405; Conservative 19; Mismatches 527; Indels 319; Gaps 48;

Qy 1 GATCTGGGTGTGCCAGGAAAAAGCAATCTGGA--AGTTAATGGTTTC--AGTGAT 56

GN 156G.
 OS Parametium primaurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Parametium.
 OX NCBI_taxID=5886;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=156;
 RX MEDLINE=87060934; PubMed=7383679;
 RA Prat A., Katinka M., Caron F., Meyer E.;
 RT "Nucleotide sequence of the Parametium primaurelia G surface protein.
 RL A huge protein with a highly periodic structure.";
 RL J. Mol. Biol. 189:47-60(1996).
 CC -!- FUNCTION: This protein is the surface antigen or immobilization
 CC antigen of Parametium primaurelia.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- DOMAIN: It has internal homologies and a highly periodic structure
 CC with 37 periods of about 75 residues, each period containing 8
 CC cysteines, except for four half periods. A variable part of 475
 CC residues comprises 4 almost identical periods in the middle of the
 CC protein.
 CC -!- MISCELLANEOUS: Expression of G protein occurs at low temperatures
 CC (14-32 degrees Celsius).
 CC -!- SIMILARITY: Contains 34 PSA repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; X03882; CAA27514.1; -.
 DR PIR; A23475; A23475.
 DR InterPro; IPR002895; Parametium SA.
 DR Pfam; PF01508; Parametium_SA; 34.
 DR SMART; SM00639; PSA; 33.
 KW Signal; Repeat; Antigen; Membrane; GPI-anchor.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2715 G SURFACE PROTEIN, ALLELIC FORM 156.
 FT DOMAIN 106 2560 37 X 75 AA APPROXIMATE REPEATS.
 SQ SEQUENCE 2715 AA; 279551 MW; 97BE359A93C7C298 CRC64;
 Query Match 11.8%; Score 893.5; DB 1; Length 2715;
 Best Local Similarity 24.6%; Pred. No. 1.3e-38;
 Matches 453; Conservative 56; Mismatches 630; Indels 701; Gaps 102;
 QY 3 TCTGGGGTGTGCGCAGGAAAAGCAAATCTGGAAGTTAATGTT-----TTTGAG 52
 DB 835 TCTVYKVGAGCC-----VKSANGSDYMTSAQCHKTLTNLTANDCKWIVDRCYALS9FA 890
 QY 53 TGATTTTAAATCC---TTGCTGGCGAGAGCGCGCTCTCCCGGTATCAGCGCT--T 107
 DB 891 TGACTTFTGTMCEYRAGCTNTVGAASSASCTLDCTLT-----GSLTFAD 939
 QY 108 CCTCATTTTGAATCCGGGCTCCGGGCTTCCTGGCGTCAGACGAGCGGAGGAGCT 167
 DB 940 CQALDSTCSVKDGT-----GCTAQSTCAGYSTAANCPR---SSASGTAGYC- 985
 QY 168 GTTTGCAATTAAGCG--GGCTGTG-----AAGCCCGCAGGCC-----CGCGGGGG 211
 DB 986 AMNTNCQSVTSAEACAFVTLGLDHSKCOLYHSSCTSLKDGTCGCEYKTTCSGYAATNN 1045
 QY 212 C---GGGCGCAGGC--GGGCCATTTTGAATAAGA--GGC-----GTGC-----CTTCCAG 256
 DB 1046 CATSGGCKFFDVECLRFNSCASITGTLTATCGTYDAGCVANVNGTACOEKLTCDLY 1105
 QY 257 GCAGGCTCTAAGTGACC-----GCCCGCGGAGCGTGGCGGTTGCAGGTCACCT---- 308
 DB 1106 LTQNSCSISAAATADKCAWSGTAFLAVTVGTHCPYVGTGLTDLICAAYNANTANKA 1165
 QY 309 GTAGCGGACTTCTTTGTTGTTTCTCTCTTTCTGGGGACCTCTGGAC-----TCACTCCCC 364

Db 1166 GTA-COEKATCNLYT--TEATCSISAAATADK-----CAWSGAACLAVTTVTECAVY 1217
 QY 365 AGATG-----AAGCGGTG--AGCCCGTGGCGGTGCTAGAGGCGGTGCTGCC 416
 Db 1218 TG--TGLTDLICAAYNANTANKAGTACOEKATCNLYTTEATCTSSAAATADK---- 1271
 QY 417 TGTGGAACGCACTCTGGCATCGCCCGGGCGCGAGGAAGCGCGCGAGCTGAGGAGC 476
 Db 1272 --AWSGAACLAVTTVTECAVYVGTGLTNAIC-----AAYNANTANKAGTACOEKAT 1323
 QY 477 CG--CTGAGTCTGTCGAGACA--TGAACCA-----C-----TGCTACTCC--GCCT 519
 Db 1324 CNLYTTEATCSTSAATAADKCAWSGAACLAVTTVTECAVYVGTGLTNAICAAYNANT 1383
 QY 520 GCGGAACTGCTACC-----CGGAGTCCGAGAGGCACTCAGC-----TTAGCCA--G 565
 Db 1384 ANRAGTACOEKATCNLYTTEATC-----STSAATAADKCAWSGAACLAVTTVTECAVY 1440
 QY 566 GTGGAATCCTACAGC-----GGTCACTGACTACATT----- 598
 Db 1441 GTGLTTRACATYAGGGINLKDGTGCGEAFKANKDYTTSNKTAQTTSLSLWINDSCYP 1500
 QY 599 -----CT-----CGACTGCAGGTATCTCTG--GCGG--AGCCAGC-----C 631
 Db 1501 VTDLNCSTVITGLGFVHAQCOAYSTGCTSVDSGSKQDPFKSTCEQVPGTTLGCTKTATKC 1560
 QY 632 CCTGGAC-----CCCTGATG-----GC-----CC-----C--CA 654
 Db 1561 YLOGSACITISNVAIDCAKITSGATITFEICQSVNTGCSVNRARSACVQQAQCSGVT 1620
 QY 655 CCTTC-----CCA-----TCAGACAGCC-----GAG-----CT--CGCTCCGAA 688
 Db 1621 AMTSCYKSGAGLCIATNTDTACVAATAATDAVILGAGNVSSANCNEMKAGCTNNGTTA 1680
 QY 689 CTGTGTCATC-----TCAACGACAAAGAGGCTTTTGCCACTGACTC-----GGCG 735
 Db 1681 CVAKTCANAGITPNHTNCNLYNTCTVNSGSACTWASKADDTQASCLYSVEGCV 1740
 QY 736 TGTCTCT-----GAC-----ACCTCCAGAAC--GCA 758
 Db 1741 VGTSCVRKTCDTAATDATRDBDTECSTYQOSTVARLGAQCARACATYKSLQCKFNTS 1800
 QY 759 GG-----TGCTGGCGCGCGCTT-----AAGCCGAGC-----GGC----- 818
 Db 1801 GSKCFWNPINKTCVDLNCNIEATLYDTHNECVAVDATLACTVATNGAAQCGMARGA 1860
 QY 794 CCTCT--CCTGCCG-----AAGCCGAGC-----GGC----- 818
 Db 1861 CASYTTIEOCKTNASNGVCVWNTNANLPAPACODKSCTSAPTSTTTHNDYAYNTATVK 1920
 QY 819 -----AGGATGGGCCCCAACTT-----CG-----CCTGC----- 844
 Db 1921 CTVATPNSGGNPTLGGCQQTAA CSSYDKQCOINANGDPGWNQTQCADKSCATASA 1980
 QY 845 -----CCTTGACT-----TC-----ACCAATCCCT 867
 Db 1981 TADYDDTKRAYITNKCTVSDSGGCVBEIPATCEMTQKCYNKNAGDPCTWTGTACIT 2040
 QY 868 TCCTGGAGA-CTAAACCTGGTCT-----C-----AGGAGCGAAGGACTGTGAAC 911
 Db 2041 KSCDNAPDATADECNVTLAGCTLNNVKCKTKVCEDFAFATDALCKQAICTCTINGNC 2100
 QY 912 TT-GT-----AGCTGAAGAC-----CAGAGC--TAGCTCTGCC 944
 Db 2101 VTGTCFQALSQAGCVTSTNQCEWIPAVLNASNVITSPAYCTIKNGSTAPITLSE-- 2158
 QY 945 ACCAG-----CT---GGCGAGCTCACCTGTCCGCCCCACC-----CCCAA 985
 Db 2159 AACAGFTNCTTKNGGCVTKSTCSAVTIDVACTTALNGTVCWDSAKNCRDRDQDFS 2218
 QY 986 GTT-----CTA-AGGTCTTT-----TCAGAGC--GTGG-----AGGTGGA 1019

| | | |
|------|--|------|
| 2219 | GTTHAACQAPAGCTAGAGGKCARVQCQTSVRACIEGTNGPCLMDIKYQNTDGTGKA | 2278 |
| 1020 | A-----GGAGT-----GGC----- | 1028 |
| 2279 | CFRYTSCSLNWNNDSSCKWISNKCTTNGSNCGVITLQSEWTNDGGCVGYDGCACIQSP | 2338 |
| 1029 | -----TGCT-----CTCCAACT-----ATGCCAAGGGCGC-----GGC----- | 1057 |
| 2339 | DLNSDPKVCKPYTSCADAFYTHTHSDCQIASSKCTTNGTGTGCIAGLGSCTSVTVQAGCYFN | 2398 |
| 1058 | -AGAGCTGGTCTTCTGGTCTCCCTTGAGAAAGGTTCTGTGGCCCTGATTTATGAACCTC-- | 1114 |
| 2399 | DKGLTYTSGVITST--GICTWDTSTSSCRDQSCADLTGTTHATCSSQLSTCTSDGTTCLL | 2456 |
| 1115 | -----TATAATAGAGT-----ATAT-----AGGTTT----- | 1135 |
| 2457 | KGACTSVYTFQCTTAVGSDGACWELASATNNTAKRLLTCADIQNGTATNVCVSVALS | 2516 |
| 1136 | -----TGTTAC-----CTT-TTTTAC-----AGAGGTGACT--TTCT--- | 1165 |
| 2517 | TCVSNMTHACIPKANCSTYTSKVACNCGGLDGCIVFTQSTATGAAGTGTCTALMTACTIVAN | 2576 |
| 1166 | -GTAAACA-----TGCGATCATATATAAACTTTTATAAA | 1199 |
| 2577 | NDQTAQAAARDCSWTAASTGRTATAVASKATHCATATNOA | 2616 |

RESULT 4

```

GI68_FARP2P      STANDARD;          PRT;   2704 AA.
ID    GI68_FARP2P
AC    P17053;
DT    01-AUG-1990 (Rel. 15, Created)
DD    01-AUG-1990 (Rel. 15, Last sequence update)
DE    10-OCT-2003 (Rel. 42, Last annotation update)
DE    G surface protein, allelic form 168 precursor.
GN    168G.
OS    Paramacium primaurelia.
OC    Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC    Paramacium.
OX    NCBI_TaxID=5886;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=168;
RC    MEDLINE=90172419; PubMed=2308165;
RA    Prati A.;
RT    "Conserved sequences flank variable tandem repeats in two alleles of
RT    the G surface protein of Paramacium primaurelia.";
RL    J. Mol. Biol. 211:521-535 (1990).
CC    -!- FUNCTION: This protein is the surface antigen or immobilization
CC    antigen of Paramacium primaurelia.
CC    -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC    -!- DOMAIN: It has internal homologies and a highly periodic structure
CC    with 37 periods of about 75 residues, each period containing 8
CC    cysteines, except for four half periods. A variable part of 475
CC    residues comprises 4 almost identical periods in the middle of the
CC    protein.
CC    -!- MISCELLANEOUS: Expression of G protein occurs at low temperatures
CC    (14-32 degrees Celsius).
CC    -!- SIMILARITY: Contains 34 PSA repeats.
CC    -----
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC    or send an email to license@isb-sib.ch).
CC    -----
DR    EMBL; X52133; CAA36378.1; -.
DR    F01; S09118; S09118.
DR    InterPro; IPR002895; Paramacium_SA.
DR    Pfam; PF01508; Paramacium_SA; 34.
DR    SMART; SM00639; PSA; 33.

```

Db 1669 TACVAKTCANAVVFNHNTCNGLNCTVNSGACQTMASKCADQTOASCLYSVEGCV 1728
QY 735 GTGTC-----TGA-----CAC-----CT-----CCAGAAGC-----CAGGT 761
Db 1729 VVGTSVRRKCTAATDATATRDDTFCASVQSQCTVARLGAQARAA CASYKSSLCQKXNT 1788
QY 762 GCTGG-----CCCCCGTT-----CTGCTGGGACCCCGG 790
Db 1789 --SGRCFWNPNTKCTVDLNCNIEASTLYDTHNECVVVDATLACTVATNGAAVQGMA 1846
QY 791 GAACCTCT-----CTGCGCG-----AAGCGGAC-----GGC-----818
Db 1847 RGACSSYTBEBCKNAGVGVCMWNTANLPAQDKSCTSAPTSTTHDNCVAYNTA 1906
QY 819 -----AGGATGGGCCCCAATT-----CG-----CCTGTC-- 844
Db 1907 TVKCTVAPNSGGNPTLGGQQTAACSSYIDKEQCOINANGBPCWNGTQCADKSCAT 1966
QY 845 -----CCACTTGACT-----TC-----ACCAATC 864
Db 1967 APATADYDDTKRAYITNKCTVSDSGQGVCEIPATCETMTQKQYVNAKADPCVWTGTA 2026
QY 865 CCTTCTGGAGA-CTAAACTGTGCT-----C-----AGGAGCGAAGGACTGTG 908
Db 2027 CITKSCDNAPDATADECNCTVLAGCTLDNVKCTKVCEDFATDALCKQAISTCTTNG 2086
QY 909 AACTT-GT-----AGCCTGAAGAC-----CAGAGC-TAGCTCTG 941
Db 2087 TNCVTRGTCFQALSQAGCVTSSTNQCEWIPAVLNASVITSPAYCTIKNCSTABITLTS 2146
QY 942 GCCACAG--CT--GGCGAGCTCACCCTGCTCCACCCACC--CC 982
Db 2147 E-AAAGFTNCTTNGGCVTKSTCSAVTIDVACTTALNGTVCAWDSQAKCRDXCQ 2204
QY 983 CRAGTT-----CTA-AGTCCTTT-----TCAGAGC--GTGG-----AGGTGT 1016
Db 2205 DFGTTHAACQAGTAGCTAGAGKCARVQNCQTSVRAACIEBTNGPCLWIDKYQNTDGT 2264
QY 1017 GGAA-----GGAGT-----GGC-- 1028
Db 2265 XGACFRYTSCKSLNWNDSCKWISNKTNGNCGVILCSETNWDGCVTGYDGACIQ 2324
QY 1029 -----TGCT-----CTCAGAACT--ATGCCAAGCGCGGCAGA 1060
Db 2325 SVPALNSSDPKCVKPYTSCADAFVTHSDQIASCKTNGTGTGCIALGSCSSY--TAQA 2382
QY 1061 GC-----TGTCTCTGCTCTCTGGAGAAAGTTCTGTCCTCGATTATGAA 1111
Db 2383 GCYFNDXGTLVTSVITST--GICTWDTSSCRDQSCADLTGTTATSSQLSTCTSDG 2440
QY 1112 CTC-----TATAATAGAT-----ATAT-----AGGTTT-- 1135
Db 2441 TTCLLKACTSYTTQTACTTAVGSDGACWELASATNNNTAKRLLTCAQIQNGTATNVC 2500
QY 1136 -----TTTAC-----CTTTT-----TTACAGGAAGGTGACT--TT 1163
Db 2501 SVALSTCVSNGTACIPKANCSTYTSKIACNSGGLDGICVFTQSTATGAAAGTGCALMTA 2560
QY 1164 CT-----GTAAACA-----TGGATGTATATAAATTTTATAAA 1199
Db 2561 CTVANNDQTACQARDRCSTWAAASGTGATAVASKATHTCATNOA 2605
RESULT 5
WA22 MYCTU STANDARD; PRT; 914 AA.
AC O06794;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE WAG22 antigen precursor.
GN WAG22 OR WAG22B OR RV1759C OR MT1807 OR MTCY28.25C OR MB1789C.

OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37Rv;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."; Nature 393:537-544 (1998).
RL [2]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RC MEDLINE=2206494; PubMed=12218036;
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains"; J. Bacteriol. 184:5479-5490 (2002).
RL [3]
RN SEQUENCE FROM N.A.
RP SPECIES=M.bovis; STRAIN=AF2122/97;
RC MEDLINE=22709107; PubMed=12788972;
RX Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Gronin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Akin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -|- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGSR
CC SUBFAMILY.
CC -|- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to a
CC frameshift in position 85.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 295890; CAB09322.1; -
DR EMBL; AE007040; -; NOT ANNOTATED CDS.
DR EMBL; BX248340; CAD94491.1; ALT_FRAME.
DR PIR; H70987; H70987.
DR TIGR; MT1807; -
DR TubercuList; RV1759C; -
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
KW Antigen; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30
FT CHAIN 31 914 WAG22 ANTIGEN.
FT SEQUENCE 914 AA; 74354 MW; F6953C3DBE856AC8 CRC64;
SQ
Query Match 10.4%; Score 784.5; DB 1; Length 914;
Best Local Similarity 29.5%; Pred. No. 1.9e-33;
Matches 337; Conservative 18; Mismatches 386; Indels 401; Gaps 48;
QY 1 GATCTGGGGTGTGTCGCGAG-----GAAAAAGCAATATCTCGGAAGTTAATGTTTTCAGTGA 55

Db 116 GNCANGAPGTGANGDGLWGLNGGAGGAGKA---NGGAGPFGAAGLFGNGAG-GA 171
 QY 56 TTTTAAATCCTTGTCTGGCGAGAGCCCGCTCTCCCGGTATCAGCGCTTCTCTCATTC 115
 Db 172 GGTATA-----NNGIGGAGG-----AGGSAMLFGAG----- 197
 QY 116 TTTGAATCCGCGCTCCGCGTCTTCCGCGTCTAGACCCAG-----CCGAGAGAGCCT 167
 Db 198 ---GAGGAGGAATSLVGGITGGTGGNAGLAGAGAGAGGAGGFSFSTAGAGAGAG--- 251
 QY 168 GTTTCGAATTTAAGCGCGCTGTGAACGCCAGG--GCCGCGCGCGCGCGCGCGCGCGG 225
 Db 252 ---AGGLFTTGGVGGAGGGGHTGGAGGAGGAGLFG-AGGAGGAGGAGGAGGAGG 307
 QY 226 GCAATTTTGAATAAGAGCGCTGCTTCCAGCAGGCTCTATAGTACCGCGCGCGCGGA 285
 Db 308 -----DGGGGLFGAG-----DGGAGGSLTGTGAAGNAGNAGTSLGAG- 349
 QY 286 GGTGCGCGCTGTGAGGTCACTGTAGCGGACTTCTTTTGTCTTCTTCTTCTTGGGCG 345
 Db 350 -----GAGGTGAGGTVFGGKGAGGA-----GGN- 375
 QY 346 ACTCTGAGTCACTCCCGAGCATGAAGCGCTGAGCCGCTGCGCGCTGCT--ACGAG 403
 Db 376 -----AGMLFGSGGGGGTGGFGAAGGGG-GVGGSGAGMLSGSGGG 415
 QY 404 GCGGTGTGCTGCTGTTCGA--ACGAGTCTTGGCCATCGCCCGCGCGCGCGGAA--G 457
 Db 416 GAGGSG-GPAGTAAGAGGAGGAGPGLIG-NGGNGGNGGSGGTGGVGGAGNAVLIENG 472
 QY 458 GCGCGCGAGCTGAGAG--CGCTGAGCTTGC-----TGGACGACAT 498
 Db 473 GGGTGTALAGSGFGGGLLGLGADYNAPSTSPWHLQDILSFNEPTALTGRLPI 532
 QY 499 GAACCACTGCTACTCCCGCTCGCGAACTGTATCCCGAGTCCCGAGAGCGCACTCAGCT 558
 Db 533 GNGDSGTG-----TGDDGGA--GGWLPFGNGNG--GAGAAGTNGSAGGAG 574
 QY 559 TAGCCAGTGGAAATCCTACAGCGGTCTAGACTACATTCTCGACCTGAGGTAGTCT 618
 Db 575 GAGGILFTGGAG-----GAGGVGTAGA-----GAGGAGGSAFL 609
 QY 619 GGCAGAGCAGCCCTGAGCCCTGATGGCCCCCCTTCCCATCCAGACAGCCGAGCT 678
 Db 610 IGSAGTGGVGAATTGG-----VGGAGG-----NAGLLIAGAAGLG-- 645
 QY 679 CGCTCGGAACTGTCTATCTCAACAGCAAAAGGAGCT-TTTGCACTGACTCGCGCTG 737
 Db 646 ---GCGGAGTAG-----VTTGGGTGGAGLFGAGGA---GGAGGTG 683
 QY 738 TCTGACACCTCCAGAACG--CAGTGTCTGCGCGCGCTTCTGCTGGAGCCCGGGAAC 795
 Db 684 STAGGAGG---AGGAGLVAHGGTGGFGNGGSGTGGGTGGAGGPGGLYAGGS--- 734
 QY 796 TCTCTGCGGAGCGGAGCGGAGGATGGCCCCCACTTGGCCCTGCCCACTTGACTT 855
 Db 735 -----GGA-----GGHGMAGGGGVGG-----NAGSLTL 759
 QY 856 CACCAAAATCCCTTCTGAGACTAACTGTGTCTGAGGAGCGGAGGACTGTGAATGT 915
 Db 760 NA-----SGGAG-----GSGGSSLSGKAGAGAG-----GS 785
 QY 916 AGCTGAGAGCCAGCTAGCTCTGCGCACAGCTGGGAGCTGACCTGCTCCAC 975
 Db 786 ASLFTGSGAGGNGGYSING---TGG-----DGTGGAQ----- 817
 QY 976 CCAACCCCAAGTTCTAAGTCTTTTCAAGCGTGGAGTGTGGAGAGTGGGTCTCTC 1035
 Db 818 -----ITGLRSFG-GAGGAGGASDTGAGGNGGAGGAG----- 850
 QY 1036 CAAACTATCCCAAGCGGCGGAGCTGCTTCTGCTCTCTCTGAGAAAGGTTCTGT 1095

Db 851 -----LYNGGCGGAGGAGGATSGKGGAGNAVVGNG-----GNGGNAGKAGGTAGAG 899
 QY 1096 TG 1097
 Db 900 AG 901

RESULT 6
 Y278 MYCTU
 ID Y278 MYCTU STANDARD; PRT; 957 AA.
 AC P56877;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Hypothetical PE-PGRS family protein RV0278C/MT0291 precursor.
 DE Hypothetical PE-PGRS family protein RV0278C/MT0291 precursor.
 GN RV0278C OR MT0291 OR MV035.06C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.,
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=2220494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
 CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL021930; CAA17953.1;
 DR EMBL; AE006936; AAK44511.1; ALT_INIT.
 DR PIR; D70835; D70835.
 DR TIGR; MT0291;
 DR TubercuList; Rv0278C;
 DR InterPro; IPR000084; PE_region.
 DR Pfam; PF00934; PE; 1.
 KW Hypothetical protein; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 30
 FT CHAIN 31 957
 FT POTENTIAL.
 FT HYPOTHETICAL PE-PGRS FAMILY PROTEIN
 FT RV0278C/MT0291.
 FT M -> I (IN REF. 2).
 FT CONFLICT 158 163 MISSING (IN REF. 2).
 FT CONFLICT 807 807 R -> G (IN REF. 2).
 FT SEQUENCE 957 AA; 81905 MW; 71EBABD417BBA47C CRC64;
 SQ

Query Match 9.5%; Score 714.5; DB 1; Length 957;
 Best Local Similarity 30.1%; Pred. No. 6.8e-30;
 Matches 352; Conservative 19; Mismatches 406; Indels 393; Gaps 58;

QY 21 AAAAGCAAAATCTGGAAGTTAATGTTTGTAGTGAATTTTAAATCTCTGTCGGGAGAG 80
 DB 12 AAAATDLASLSSSI SAANAAAAAATTAALMAAGADEVSTAAAL-----FGAH 58
 QY 81 GCCGCGCTCTCCCGG---TATGAGCGTCTCTCTATT-----CTTTG----- 119
 DB 59 GQAYALSAQAQAFAHQVQLTSGGAYAAEAHAAVPLDPIEFNFFLANTGRPLING 118
 QY 120 --AATCCGCGCTCGCGGTCTTCCGCGTCAGACCGAGAGAAAGCTTTTTCGAATT 177
 DB 119 ANGAPETGANG---GDGWLINGGAG--GSGAGVNGGAGNGG----- 158
 QY 178 TAAGCG--GGCTGTGAACCGCCAGCGCGCGCGCGCGG---GCCGAGCGCGGCATT 231
 DB 159 -AGNGGAGGLINGGAGG---AGVASSGIGGSGAGAGNAMLFGAGGAGGAGGVA-- 212
 QY 232 TTGAATAAGAGGCGTGTCTTCCAGGAGGCT---CTATAAGTGACCGCGCGCGAGCG 288
 DB 213 LTGGAGGAGGAGG--NAGLLFGAAGVGGAGGFTNGSALGGAGGAGGLFATGGVGGSG 271
 QY 289 TCGCGCGCTTGCGAGTCACTGAGCGGACTTCTTTTGTGTTTCTTCTCTTTGGGCGACC 348
 DB 272 AGSSGAGGAGGAGGLFGAGGTGGHGFADSFSGVGG----- 309
 QY 349 TCTGACTCATCCCGACATGAAGCGCTGAGCCCGTGGCGGCTGTCTAG--AGGCGG 407
 DB 310 --AGG-----AGGLFGAGGGBGGGSHLVAGG--GAGNAGMALGAAGAGG 355
 QY 408 TGTGTGCTCT--GTCCGAACGAGCTGTGCGCATCGCCGCGCGCGG---GAAGGCGC 461
 DB 356 IG-GDGGTLTAGGIGA-----GGAGNAGLLFGSGGSGAGGFGFADGGQGP-- 403
 QY 462 CGGAGCT--GAGGAGCGCTGAGCTTGC--TGAGAGCATGAACCACTGTCTCCCGCC 518
 DB 404 -GGNAGTVFGSGAGGNGVGGVGGFAGGAGGAGTGGPGLINGGN----- 445
 QY 519 TCGGGAAGTGTGTAACCGAGTCCGAGAGGCACTCAGCTTAGCCAG---GTGGAATCC 575
 DB 446 GNGGASAVTGGG--GIGGTGVLINGNGSGGIGAGKAGVGVSGLLGLDGFNPAS 504
 QY 576 TA-----CAG---CGGTCTGACTACATTCGACCTGAGGTA 613
 DB 505 TSPHTLOQNVLVNVPFQTLGRPLINGANGTPG-----TGADGGAG 549
 QY 614 GTCTGCGCGAGCCAGCCCTGAGACCCCTGATGGCCCGCCACCTCCCATCCAGACAGCC 673
 DB 550 GWLFGNGANG-----TPG-----TGAAGG-----AGGWLFGNG 577
 QY 674 GAGCTCGCTCGGACTTGTCTCATCTCCAAAGAGAGCTTTTCCACTGACTCGGC 733
 DB 578 GNG-----GHGA--TNTAATAT-----GGAGGAG-----GILFGT----- 508
 QY 734 CGTGTCTTGACACCTCCAGAACGAGGTGCTGGCGCGCTTCTGCTGGGACCCCGGAA 793
 DB 609 GTGGIATGA-----GGGAGGAGGVVLLIGSGGTG--GNGGNSIGVAGIGA 655
 QY 794 CTTCTCTCCCGGAGCGGAGC---GCAG-----GGATGGCCCACTTCGCCCTG 843
 DB 656 -----GGGGDAGLLFGAAGTGGHGAAGVPAVGAGGNG-----GLFANG 697
 QY 844 CCACTTGACTTCAACAAATCCCTTCTGGAGACTAACTGCTGTCTGAGGAGCGAAGA 903
 DB 698 -----GA-----GGAGGFNA-----GGNGN--GGLFTGTGTGA 726
 QY 904 CTGTGAATGTAGCTGAAGCCAGAGCTAGCTCTGGCCACAGCTGGCGCACTCAC 963
 DB 727 GTFAG-----GNGGNGGLFGAGGTGGAAGSG-----SGITTTGGGHHGNAGL 771

964 CTTGCTCCACCCACCCCAAGTTCTTAAGGTTCTTTTTCAGAGCGTGGAGGTGGA---- 1019
 772 LSLG-----ASG-----GAGG-----CGASSLAGAGGTGGGALLFG 805
 1020 ---AGGAGTGGCTCTTCAAACTATGCCAAGCGGCG---GCAGAGCTGTCTTC-- 1070
 806 FRGAGGAGHG-----GAALTSIQGAGGAGGGLLFGSAGAGGAGSGANALG 856
 1071 --TGGTCTCTCGGAGAAAGGTTCTCTTGC 1098
 857 AGTGG-----TGGDSGHAGVFGNGDGGC 880

RESULT 7
 ID NTC1 RAT STANDARD; PRT; 2531 AA.
 AC Q07008;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
 GN NOTCH1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Schwann cell;
 RX MEDLINE=92111383; PubMed=1764995;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "A homolog of Drosophila Notch expressed during mammalian
 RT development.";
 RL Development 113:199-205 (1991).
 RN [2]
 RP REVISIONS TO 1652-1653.
 RA Weinmaster G.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RX MEDLINE=21094508; PubMed=11182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Honjo T.;
 RT "Notch1 and Notch3 instructively restrict BGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate.";
 RL Neuron 29:45-55 (2001).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941 (1992).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181 (2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(SC) which are probably linked by disulfide
 CC bonds (By similarity).

| | | | |
|------|----|---|------|
| 857 | Qy | ACCAATCC-----CTTC-----CTGAGACTAAACTGCT-----GCTCAGG | 894 |
| 991 | Db | SCFNGGTCVNDGINSFTCLCPPGFTGSYQYDVNEDSRPCLHGHTCQDSYGVKCTCPOG | 1050 |
| 895 | Qy | AGCGAA-----GGACTGTG-----AACTTGTAG----- | 917 |
| 1051 | Db | YTGILNCQLVRWCDSPCKNGGKWCQTNTQYHCECRSGWTGFCNDVLSVCEVAAAKRGI | 1110 |
| 918 | Qy | -----CCTGAAGAGC-----CAGAGCTAGCTCT-----GSCCA | 945 |
| 1111 | Db | DVTLQHQHGLCVDEBDKHCHYCHQAGYGSYCEBDEVDCSPNQCNGATCTDYLGGFSCK | 1170 |
| 946 | Qy | CCAGCTGGGGC-ACGTAC--CCTGCTCCACCCACCC----- | 981 |
| 1171 | Db | CVAGYHGSNCSEINECLSQPCQNGTCTDLTNTYKSCPRGTQGVHCEINVDDCHPPLD | 1230 |
| 982 | Qy | -----CCAGTTCCTAAGG-TCTTTTCAGACGCTGGAGGTGTGGAAGGAGTGGCTGCTC | 1033 |
| 1231 | Db | PASBPCKFFNGFVCDVQGVYCT-----CPPGFVGER-----CEGDVN | 1269 |
| 1034 | Qy | TCCAAACTATG-----C-CAAGGGCG-----CGGCAGAGC-TGGTC-----TTC | 1070 |
| 1270 | Db | ECLSNPCDPRGTQCVORYNDFHCECRAGHTGRRCESVINGCRKPCRNQGVCAVASNTA | 1329 |
| 1071 | Qy | TGGTCTC-----C-----TTGAGAAAGGTTCTG--TTGCCCTCATTTATGAAGTCT | 1115 |
| 1330 | Db | RGFICRPFARFEGATCENDARTCGSLRCLNGTCTISGRSPSTCULG-----SFTGPECQFP | 1386 |
| 1116 | Qy | ATAATAGA-----GTATATA | 1130 |
| 1387 | Db | AS8PCVGSNFCYNQGTCEPTS | 1407 |

RESULT 8

NTCL_HUMAN

ID

NTCL_HUMAN

STANDARD;

PRT; 2556 AA.

AC

P46531;

01-NOV-1995 (Rel. 32, Created)

28-FEB-2003 (Rel. 41, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)

Neurogenic locus notch homolog protein 1 precursor (Notch 1) (hnl)

(Translocation-associated notch protein TAN-1).

NOTCH1 OR TAN1.

Homo sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxId=9606;

[1]

SEQUENCE FROM N.A.

TISSUE=Brain;

Mann R.S., Blaumueller C.M., Zagouras P.;

"Complete human notch 1 (hnl) cDNA sequence.";

Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE OF 1-2444 FROM N.A.

MEDLINE=91347367; PubMed=1831692;

Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,

Smith S.D., Sklar J.;

"TAN-1, the human homolog of the Drosophila notch gene, is broken by

chromosomal translocations in T lymphoblastic neoplasms.";

Cell 66:649-661(1991).

[3]

IDENTIFICATION OF LIGANDS.

MEDLINE=99180765; PubMed=10079256;

Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Cartangu M.-L.,

Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;

"Human ligands of the Notch receptor.";

Am. J. Pathol. 154:785-794(1999).

[4]

INTERACTION WITH DTX1.

MEDLINE=98250176; PubMed=9590294;

Matsuno K., Eastman D., Mitsiadis T., Quinn A.M., Cartanciu M.L.,

Ordentlich P., Kadesch T., Artavanis-Tsakonas S.;

RT "Human deltex is a conserved regulator of Notch signalling.";
 RL Nat. Genet. 19:74-78(1998).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May be important for normal lymphocyte
 CC function. In altered form, may contribute to transformation or
 CC progression in some T-cell neoplasms. Involved in the maturation
 CC of both CD4+ and CD8+ cells in the thymus (By similarity).
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(IEC) which are probably linked by disulfide
 CC bonds (By similarity). Interacts with DTX1 and DTX2.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -!- TISSUE SPECIFICITY: In fetal tissues most abundant in spleen,
 CC brain stem and lung. Also present in most adult tissues where it
 CC is found mainly in lymphoid tissues.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- DISEASE: NOTCH1 truncation is associated with T-cell acute
 CC lymphoblastic leukemia.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC -----
 CC EMBL, AF308602; AAG33848.1; --
 CC EMBL, M73380; AAG6014.1; --
 CC HSPF, F00740; IEDM.
 CC Genew; HGNC:7881; NOTCH1.
 CC MIM; 190198; --
 CC GO; GO:0015021; C:integral to membrane; NAS.
 CC GO; GO:0003793; F:defense/immunity protein activity; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_s.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000500; Notch_dom.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00008; EGF; 5.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFELOOD.
 CC PRINTS; PR00011; EGFAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.

DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 20.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 26.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 18.
 KW Receptor; transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 18
 FT CHAIN 19 2556
 FT CHAIN 1722 2556
 FT
 FT CHAIN 1755 2556
 FT
 FT
 FT DOMAIN 19 1736
 FT TRANSMEM 1737 1757
 FT DOMAIN 1758 2556
 FT DOMAIN 20 58
 FT DOMAIN 59 99
 FT DOMAIN 102 139
 FT DOMAIN 140 176
 FT DOMAIN 178 216
 FT DOMAIN 218 255
 FT DOMAIN 257 293
 FT DOMAIN 295 333
 FT DOMAIN 335 371
 FT DOMAIN 372 410
 FT DOMAIN 412 450
 FT DOMAIN 452 488
 FT DOMAIN 490 526
 FT DOMAIN 528 564
 FT DOMAIN 566 601
 FT DOMAIN 603 639
 FT DOMAIN 641 676
 FT DOMAIN 678 714
 FT DOMAIN 716 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 868
 FT DOMAIN 870 906
 FT DOMAIN 908 944
 FT DOMAIN 946 982
 FT DOMAIN 984 1020
 FT DOMAIN 1022 1058
 FT DOMAIN 1060 1096
 FT DOMAIN 1098 1144
 FT DOMAIN 1146 1182
 FT DOMAIN 1184 1220
 FT DOMAIN 1222 1266
 FT DOMAIN 1268 1306
 FT DOMAIN 1308 1347
 FT DOMAIN 1349 1385
 FT DOMAIN 1388 1427
 FT REPEAT 1446 1481
 FT REPEAT 1482 1523
 FT REPEAT 1524 1563
 FT REPEAT 1928 1957
 FT REPEAT 1961 1991
 FT REPEAT 1995 2024
 FT REPEAT 2028 2057
 FT REPEAT 2061 2090
 FT DOMAIN 1576 1665
 FT DOMAIN 1662 1699
 FT DOMAIN 1729 1732
 FT DOMAIN 1741 1744
 FT DOMAIN 1902 1905
 FT DOMAIN 2260 2263
 FT DOMAIN 2404 2407
 FT DOMAIN 2411 2418
 POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 6.
 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 10.
 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 20.
 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 22.
 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 24.
 EGF-LIKE 25.
 EGF-LIKE 26.
 EGF-LIKE 27.
 EGF-LIKE 28.
 EGF-LIKE 29.
 EGF-LIKE 30.
 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 33.
 EGF-LIKE 34.
 EGF-LIKE 35.
 EGF-LIKE 36.
 LIN/NOTCH 1.
 LIN/NOTCH 2.
 LIN/NOTCH 3.
 ANK 1.
 ANK 2.
 ANK 3.
 ANK 4.
 ANK 5.
 POLY-VAL.
 POLY-ARG.
 POLY-PRO.
 POLY-ALA.
 POLY-GLU.
 POLY-GLY.
 POLY-GLN.
 POLY-PRO.

| FT | DOMAIN | 2522 | 2525 | POLY-SER. |
|----|----------|------|------|--|
| FT | SITE | 1665 | 1666 | CLEAVAGE (BY FURIN-LIKE PROTEASE) (BY SIMILARITY). |
| FT | DISULFID | 24 | 37 | BY SIMILARITY. |
| FT | DISULFID | 31 | 46 | BY SIMILARITY. |
| FT | DISULFID | 48 | 57 | BY SIMILARITY. |
| FT | DISULFID | 63 | 74 | BY SIMILARITY. |
| FT | DISULFID | 68 | 87 | BY SIMILARITY. |
| FT | DISULFID | 89 | 98 | BY SIMILARITY. |
| FT | DISULFID | 106 | 117 | BY SIMILARITY. |
| FT | DISULFID | 111 | 127 | BY SIMILARITY. |
| FT | DISULFID | 129 | 138 | BY SIMILARITY. |
| FT | DISULFID | 144 | 155 | BY SIMILARITY. |
| FT | DISULFID | 149 | 164 | BY SIMILARITY. |
| FT | DISULFID | 166 | 175 | BY SIMILARITY. |
| FT | DISULFID | 182 | 195 | BY SIMILARITY. |

Query Match 8.9%; Score 674.5; DB 1; Length 2556;
 Best Local Similarity 23.5%; Pred. No. 3.5e-27;
 Matches 374; Conservative 33; Mismatches 633; Indels 549; Gaps 93

| | | | | | | |
|----|-----|---|---|--|--------------------|-----|
| Qy | 1 | GATCTGGG---- | GTGCTGCCAGAAAAAGCAAAATTC | TGGAAGTAATGTTGTTT | GAGT--- | 53 |
| Db | 28 | GETCLNGKCEAANGTEACVCGGAVGPRCQDPNPCL | ----- | GATTTTAAATCTCTCTGCGGAGAGCCGCTCT | ----- | 76 |
| Qy | 54 | ----- | ----- | GATTTTAAATCTCTCTGCGGAGAGCCGCTCT | ----- | 92 |
| Db | 77 | VDRGVADYACSCALGFSPLCLPLDNACLTNFC | ----- | RVGTCDLLLTBYKCR | 128 | |
| Qy | 93 | CCGGTATCAGCGTTCCTCATCTTTTGAATCCGGGGT | ----- | CCGC----- | GGTC-TT | 140 |
| Db | 129 | CPFG----- | WSGKCCQADPCASN----- | PCANGGQCLPFEASYICHCPSPFHGPTCRD | 178 | |
| Qy | 141 | CGCGCGTCAGAC-CAGCGCGAGGAAGCCTGTTGCAATTTAAGC | ----- | ----- | GGG | 185 |
| Db | 179 | VNECGKPLRCHGGTCHNEVGSYRCV----- | CRATHGTGNCERPVVPCSPSQNGT | 232 | | |
| Qy | 186 | CTGTG-AAGCCCGAGCGCGGG | ----- | GGGGCGGGCCGAGCGGGGC-CATTTTGAAT | 237 | |
| Db | 233 | CRPTGVTHCEACLPFTGQNCBENIDPCGNCKNGGACV | VDGVNTVNCPCPEPWGQ | 290 | | |
| Qy | 238 | AAAGAGCGTGCGCT----- | TCAGGCA----- | GGCTCTATAAGTAGCAGCG----- | CC | 278 |
| Db | 291 | ----- | YCTBDVDECOLMPNACQNGGTCHNTHGGYNCVNVGWTGEDCSENIDDC | 339 | | |
| Qy | 279 | GCGGC--GAGC----- | GTGCGCGCGTTCACGTCACGTAGCGGACCTCTTTTGGTTTTC | 331 | | |
| Db | 340 | ASNA CFHGTCHDRVASFYCEPHRTG----- | ----- | LLCHLNDACISNCP | 381 | |
| Qy | 332 | TTTCTCTTTGGGGCACTTC----- | TGGACTACTCCCGCATGGAAGCGCTGAGCGCGGT | 387 | | |
| Db | 382 | NEGSNDTNFVNGKAICTCPSGYTGAPCSQDVDC----- | ----- | SLGANCEHAGKINTPL | 433 | |
| Qy | 388 | G----- | CGC-GGC-TG----- | CTACAGAGGCGGTGCTGCTGTGCGAAACGAGCTGGCCA | 437 | |
| Db | 434 | GSFECCLQGYTGPRCEIDVNECVSNPONDAT----- | CLDQIGEFCMCFEGYGVHCE | 488 | | |
| Qy | 438 | ----- | TCGC----- | CGAGGGAAGGC--CGGGAGCTGAGGACGC | 478 | |
| Db | 489 | VNTDECASSPLHNGRCLDKINEFCECPTGFTGHLQYDVEDCASTPCKNAGAKLDGPN | 548 | | | |
| Qy | 479 | ----- | CTGAGGTCTGTGGAC----- | GACATGAA--CCAC----- | 507 | |
| Db | 549 | TVTCV--- | CTEGYTGTHCEVIDECDPDPCHYGSCKDGVATFTCLCRPGYTGHHCEFTIN | 605 | | |
| Qy | 508 | CTACTC----- | ----- | CGCCTCGGG----- | GAACGTGATCCCGGATCC | 542 |
| Db | 606 | ECSSQCFRLRGTCQDPDNYLFCFLKGTGTGNCBENILDDCASSPDCSDGTCCLKIDGYECA | 665 | | | |
| Qy | 543 | CGAGAGGCACT----- | CAG--CTTAGCCAGGTGAAA----- | TCCTACAGCGCGTC | 586 | |
| Db | 666 | CEPGYTGSCMNIDECAGNPCHNGGTCEDINGFTCRCPGEGYDPTCLSEVNECNINPC | 725 | | | |

172 GGDLVSGHAGVGGGGWLLNGGTTGGAGGAGATLVGGTGGVGGATGLIGSGGFGG-- 230
249 CTTCCAGGAGGCTCTATAAGTGAACCGCGCGGAGCGTGGCGGCTTTCAGAGTCACT 308
231 -----AGGAAGAGVGTGGVGGGCGVGGVFGNGGFG-----GAG-GLGAAGGVGAASYF 278
309 GTAGCGAGCTTCTTTTGGTTTCTTCTTCTTGGGCACTCTGGACTCACTCCAGCA 368
279 GTGGGG-----VGG----- 288
369 TGAAGCGCTGAGCGCGTGGCGGTCTACGAGCGGTGTCTGCTCTCGGAAAGCA 428
289 DGAPGGDG--GAGPLLING-GVGLGGAGA--AGNGGAG-----GMLLG--DGGAGGQ 337
429 GTCTGGCATCGCCCGGCGGAGGAGGCGCGGAGCTGAGGAGCGCTGAGTTC 488
338 GPAVAGVL-----GGMPGAGN-----GGNANWFGGAGGCGGTGAGTNGV 380
489 TGAACGACATGAACCACTGCTACTCCCGCTGGCGGAACTGGTACCGGAGTCCGAGAG 548
381 NPGSIANPNTGANGTDSG-----NGNQTGGNGGPGAGGVG--EAGGVG-----GQG 426
549 GCATCTAGCTTAGCGAGGTGGAATCTACAGCGCTCATCGACTACATCTCGACTGC 608
427 GLGESLDG-----NDGTGG-----KGG 443
609 AGGTAGTCTCTGGCGGAGCGCCCTGATGCGCCCTGATGCGCCCACTTCCCATCCAGA 668
444 AGTACT--DGGAGAG--GAGGIGETDSAGVATGEGG-----DGA 483
669 CAGCGAGTCTGCTCCGGAATCTGCTATCTCCAAAGCAAGAGAGTTTTCAGTGCAC 728
484 TGGVDG-----VGG-----AGKGGGQGHNTGVG--DAFGD-- 513
729 TCGGCGCTGCTCTGACACCTCCAGAACGAGGTGCTGGCGCCCTTCTGCTGGGACCC 788
514 --GGIGSDG-----NGALGAGNGGTGGAG-----GNGGCGMLI 547
789 GGGAACTCTCTCCCGAAGCGGAGGAGTGGCGCCCACTTGGCCCTGCGCCAC 848
548 GNGCA-----GGAGTGTGGGGAAGPAGG--VGG----- 575
849 TTGACTTCAACAATCTCTCGGAGACTAACTCTGCTCAGGAGCGAGGAGTGTG 908
576 -----AGBGLTDTAGTAEGTG--GLGGLGVGTGNGSGG 612
909 AACTTGTAGCTGAAGAGCCAGCTAGCTCTGGCCACAGCTGGGCGAGTCACTCCCTGC 968
613 VGGNGGAAGSLI GLGGGGGAGGVGTGGIGG-----AGNGG----- 652
969 TCCACCCCGCCCAAGTTCTAAGTCTTTTTCAGAGCGTGGAGGT--GTGAGGAGTGG 1027
653 -----AGGAGTTT---GGGATGGGGGTGGVGGAGGTGTGG 686
1028 CTGCTCTCCAACTATGCCAGGCGCGGCGAG--AGCTGGTCTTCTGCTCTCTTGGAGAA 1086
687 AGTGT-----GGSGAGGLIGWAGAGGTGAGGTGG-----QGGLG-G 723
1087 AGTGTCTGCTCCCTGATTTATGAATCTATATAGAGTATATAGTTT 1134
724 QGNGGNGGTG-----ATGGGGDFALGNGGAGGAGGSGGSG 762

RESULT 11
LORI_MOUSE
ID LORI_MOUSE STANDARD, PRT; 481 AA.
AC P18165;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Loricrin.
GN LOR.
OS Mus musculus (Mouse).

RP SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch K., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=1218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z80225; CAB02341.1; --
CC EMBL; AE007103; AAK47026.1; ALT_INIT.
CC EMBL; BX248343; CAD94852.1; --
CC PIR; F70963; F70963.
CC TIGR; MT2712; --
CC TubercuList; Rv2634c; --region.
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.
CC ProDom; PD001223; PE_region; 1.
CC Hypothetical protein; Complete proteome.
CC CONFLICT 51 51 V -> L (IN REF. 2).
CC CONFLICT 63 63 Q -> H (IN REF. 2).
CC CONFLICT 274 274 A -> T (IN REF. 2).
CC CONFLICT 778 AA; 63131 MW; DAB20F58E4999E7 CRC64;
CC SEQUENCE
DR Query Match 8.9%; Score 668; DB 1; Length 778;
DR Best Local Similarity 29.3%; Pred. No. 1.3e-27;
DR Matches 295; Conservative 12; Mismatches 317; Indels 384; Gaps 44;
QY 133 CGGCTCTCGGCTCAGACCGGAGGAGCGTGTTCATTTAAGCGGCTGTGAA 192
DB 133 GPGGLLLGNGGNG---GSGAPGPGGAGGAGL-----IGNG--GTGK 171
QY 193 CGCCAGCGCGCGGCGGCGG--CGGGCGGAGCGGCGC--ATTTGAATAAGAGCGGTG 248

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90275605; PubMed=2190691;
 RA Mehler T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
 RA Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinart P.M.,
 RA Yuspa S.H., Roop D.R.,
 RT Identification of a major keratinocyte cell envelope protein,
 RT "loricrin",
 RL Cell 61:1103-1112 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/c;
 RX MEDLINE=95256248; PubMed=7738016;
 RA DiSepio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
 RA Roop D.R.,
 RT "The proximal promoter of the mouse loricrin gene contains a
 RT functional AP-1 element and directs keratinocyte-specific but not
 RT differentiation-specific expression";
 RL J. Biol. Chem. 270:10792-10799 (1995).
 CC -!- FUNCTION: Major keratinocyte cell envelope protein.
 CC -!- SUBUNIT: Monomers are crosslinked by disulfide and N-(gamma-
 CC glutamyl) lysine isopeptide bonds.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M34398; AAA39444.1; -;
 DR EMBL; U09189; AAA82152.1; -;
 DR PIR; A35628; A35628.
 DR HSRP; P02876; 9WGA.
 DR MGD; MGI:96816; Lor.
 KW Keratinization.
 SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;

 Query Match 8.8%; Score 666; DB 1; Length 481;
 Best Local Similarity 29.9%; Pred. No. 1-2e-27;
 Matches 206; Conservative 48; Mismatches 190; Indels 244; Gaps 35;

 QY 147 CAGACACCGCGAGGAGCCCTGTTTCAATTTAAGCGGCGTGTGAACGCCCGCGCGG- 205
 DB 11 CFPVCGKTSGGGGGGGGGGYSG-----GGSGCGGGSSGGSSCG--GGGGGGYGGG 59
 QY 206 --CGGGGGCGGG--GCGGAGGGGGGCCCATTTTGAATAAGAGGGCGTCCCTCCAGGCA 259
 DB 60 SSGGGGGGGGKXSGGGGGSSCGG-----YSGGGG-----SSCGGGYS 101
 QY 260 GGCTCTATAGTACCGCGCGCGGCGAGCGTGTGCGGCGTGTGACGTCACTGTAGCGGACTT 319
 DB 102 GG-----GGGSSCG-----GGYSGGGGGSSCGGSSCGGGGGGGG--- 143
 QY 320 CTTTGGTTTCTTCTCTTTGGGGCACTCTGGACTCACTCCCGAGCATGAAGGGCGTGTG 379
 DB 144 -VKYSGG-----GGGGSSC---GG-----GSSGGGGGGSSCGG 173
 QY 380 AGCCGCGTCCGCGCTGTACAGGCGCGTGTGCTGCTGCGAAGCAGCTTGGCCATC 439
 DB 174 GS---GGGGSYCGSSG--GGSSGGCGGGSG--GKYSGGGGGSSCG--GGYSGG----- 219
 QY 440 GCCCGGGCCGAGGAGAGGGCCCGCGAGCTGAGGAGCGCGTGTGCTGAGCAGCATG 499
 DB 220 GSSSGSSC--GGYSGGGSSCGGGGGYSGGGTSCGG-----GSSGGGGGGSSSQ 269
 QY 500 AACCACTCTACTCCGCGCTCGGGAAGTGTGTACCGGAGTCCGAGAGCACTAGCTT 559
 DB 270 YQCSYGG-----GSSGGSSCGGGYS---GGGGSSCGGGYS---GGGGSSCGG 311

QY 560 AGCCAGGTGAATCCTACAGCGCGTCACTGACTACATTCGACCTCGAGGTAGTCCCTG 619
 DB 312 -----GSSGGG-----SSCGSGGG-----GYSGG-----GG 333
 QY 620 GCGAGCGAGCCCTGACCCCTCATGCGCCCGACCTTCCCATCCAGACAGCCGAGCTC 679
 DB 334 GSCGGG-----SSGGGGYSSQQTSTSCAPQOSYG--GSSSGGGGSC 375
 QY 680 GCTCCCGGAACCTTGTCTATCTCCACGACAAAGGAGCTTTTCCCTGACTCGCGCGTCTC 739
 DB 376 GGGSSG-----GGG-----GYSGGGGGGSSGGCGG--- 403
 QY 740 CTGACACCTCCAGAACGAGGTGCTGGCGCCCGTTCCTCGGGACCCCGGAAACCTCTC 799
 DB 404 -----YSGGGGGCGGGS--SGSGGGGGCGGSSG--GSGGG-----CGGGYS----- 440
 QY 800 CTGCCGGAAGCGGACGCGGCGGCGGATGGG 827
 DB 441 -----GGGG--GSSCGGGSSGGGGG 461

 RESULT 12
 NTC1 MOUSE
 ID NTC1 MOUSE STANDARD; PRT: 2531 AA.
 AC Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9R0X7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A)
 DE (nt14) (p300).
 GN NOTCH1 OR NOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.,
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch";
 RL Genomics 15:259-264 (1993).
 RN [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RX MEDLINE=93050801; PubMed=1426644;
 RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.,
 RT "Expression analysis of a Notch homologue in the mouse embryo";
 RL Dev. Biol. 154:377-387 (1992).
 RN [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.,
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development";
 RL Development 115:737-744 (1992).
 RN [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.,
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues";
 RL Exp. Cell Res. 204:364-372 (1993).
 RN [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE=99364499; PubMed=10437788;

RA Lee J.S., Ishimoto A., Yanagawa S.I.:
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
 RT to induction of HES-1 in a mouse T lymphoma cell line, DU-3.";
 RL FEBS Lett. 455:276-280(1999).
 RN [6]
 RP SEQUENCE OF 1950-2201 FROM N.A.
 RX MEDLINE=98029496; PubMed=9384671;
 RA Messerle M., Folle M., Nehls M., Eggert H., Buehm T.:
 RT "Dynamic changes in gene expression during in vitro differentiation of
 RT mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
 RN [7]
 RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
 RP MUTAGENESIS OF 1651-ARG--ARG-1654.
 RX MEDLINE=98318619; PubMed=9653148;
 RA Logeat F., Bessia C., Brou C., Lebail O., Jarriault S., Seidah N.G.,
 RA Israel A.:
 RT "The Notch1 receptor is cleaved constitutively by a furin-like
 RT convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
 RN [8]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.:
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [9]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Horijo T.:
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 RN [10]
 RP INTERACTION WITH DTX1 AND DTX2.
 RX MEDLINE=21123790; PubMed=11226752;
 RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
 RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Arcavanis-Teakonas S.,
 RA Okano H., Matsuno K.:
 RT "Murine homologs of deltex define a novel gene family involved in
 RT vertebrate Notch signaling and neurogenesis.";
 RL Int. J. Dev. Neurosci. 19:21-35(2001).
 CC [11]
 CC FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation. May be involved in mesoderm
 CC development, somitogenesis and neurogenesis. Involved in the
 CC maturation of both CD4+ and CD8+ cells in the thymus.
 CC [12]
 CC SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(BC) which are probably linked by disulfide
 CC bonds. Interacts with DTX1 and DTX2.
 CC [13]
 CC SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC [14]
 CC ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q01705-1; Sequences=Displayed;
 CC Name=2;
 CC IsoId=Q01705-2; Sequences=VSP_001402, VSP_001403, VSP_001404;
 CC Note=No experimental confirmation available;
 CC [15]
 CC TISSUE SPECIFICITY: Highly expressed in the brain, lung and
 CC thymus. Expressed at lower levels in the spleen, bone-marrow,
 CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
 CC muscle, kidney and heart.
 CC [16]
 CC DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
 CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
 CC endothelial cells, while much lower levels are seen in the

CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
 CC the neuroepithelium. At 13.5 dpc expressed in the surface
 CC ectoderm, eye and developing whisker follicles.
 CC [17]
 CC PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(BC). Following
 CC ligand binding, it is cleaved by TNP-alpha converting enzyme called
 CC (FACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC [18]
 CC PTM: Phosphorylated.
 CC [19]
 CC SIMILARITY: Belongs to the NOTCH family.
 CC [20]
 CC SIMILARITY: Contains 36 EGF-like domains.
 CC [21]
 CC SIMILARITY: Contains 3 Link/Notch repeats.
 CC [22]
 CC SIMILARITY: Contains 5 Ank repeats.
 CC [23]
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC [24]
 CC EMBL; Z11896; CAA77941.1; -;
 CC EMBL; L02613; AAK14898.1; -;
 CC EMBL; X68278; CAA48339.1; -;
 CC EMBL; AJ238029; CAB40733.1; -;
 CC EMBL; X82562; CAA57909.1; -;
 CC EIR; A46019; A46019; -;
 CC FIR; B49175; B49175; -;
 CC HSSP; P00740; LEDM.
 CC MGD; MGI:97363; Notch1.
 CC GO; GO:0005887; C:integral to plasma membrane; IC.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC GO; GO:0030154; P:cell differentiation; IMP.
 CC GO; GO:0007386; P:compartment specification; IMP.
 CC GO; GO:0007219; P:N signaling pathway; IC.
 CC GO; GO:0045944; P:positive regulation of transcription from P. . . ; IDA.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR001152; Asx hydroxyl_S.
 CC InterPro; IPR000742; EGF 2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000800; Notch_dom.
 CC Pfam; PF00023; ank; 7.
 CC Pfam; PF00008; EGF; 35.
 CC Pfam; PF00066; notch; 3.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFBLD.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRINTS; PR01452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 24.
 CC SMART; SM00004; NL; 2.
 CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE; PS50088; ANK_REPEAT; 2.
 CC PROSITE; PS00010; ASX_HYDROXYL; 22.
 CC PROSITE; PS00023; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 27.
 CC PROSITE; PS50026; EGF_3; 36.
 CC PROSITE; PS01187; EGF_CA; 21.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing.

FT SIGNAL 1 18
 FT CHAIN 19 2531
 FT CHAIN 1711 2531
 FT CHAIN 1744 2531
 FT DOMAIN 19 1725

POTENTIAL.
 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 NOTCH EXTRACELLULAR TRUNCATION.
 NOTCH INTRACELLULAR DOMAIN.
 EXTRACELLULAR (POTENTIAL).

Query Match 8.8%; Score 661.5; DB 1; Length 2531;
 Best Local Similarity 23.0%; Pred. No. 6.6e-27;
 Matches 352; Conservative 36; Mismatches 596; Indels 549; Gaps 85;

QY 1 GATCTGGGTGCTGCAGGAAAGCAATTCCTGGAAGTAATGTTTGAATGATTTT 60
 DB 21 GURCSOPSGT- - - - -CLNGRCEVASGTEACVAGSVFGORCDPN- - - - - 61
 QY 61 AAATCCTTGC- -TGCGGAGAGCCCGCTCTCCCGGTATCAGCGCTTCTCATCTTT 118
 DB 62 - - -PCLSTRKNAGTCVVVDHGGVINDYACSLGFSGP- - - - -LCLTFLDKP 105
 QY 119 GAATCCGGCTC- - - - -CGCGTCTTCGGCTCAGACAGCCGAGGAAGCT 167
 DB 106 CLANPCRNGG-TCBLLTLTEYKRCRPSGWSKSCQADPCANPCAN- - - - - 151
 QY 168 GTTTGCAATTAAGCGGCTGTGAAGCCCA- - - - -GGGC- - -CGCGGGG- -CGGGGCC- 218
 DB 152 - - - - -GGCLPFESSYICRPPGFHGTCTCRQDVNECSQNPGLCRHGHCH 196
 QY 219 - - -GAGCGGGCATTTTGAATAAGAGCGTGC- - -CTTCAGCGAGGCTCTATAAGTG 272
 DB 197 NEIGSYRC- -ACCATHT- - - - -GPHCELPYVPCSPSPQNGATCRPTGTH 241
 QY 273 ACCGCGGCGAGCGTGCAGGTGAGTGTGAGTCACTGTAGCGGACTTCTTT- - - 323
 DB 242 EC- - -ACLPFAGQNEB- - - - -NVDDCFGNCK- - - - -NGACVDGVNTYNCRCPEV 287
 QY 324 TGGTTT- - - - -TCTTTCTTTGGG- -GACCT-CTGAGCTACTCCCGCAGCATG 370
 DB 288 TGYCTEDVBCQLMPNACAGTCHNTHGGYVCVNGWIGEDCSENIDDCASAACTG 347
 QY 371 AAGCGCTGAGCCC- -GTCGGGGCTGCTA- - -CGAGGCGTGTGCTGCTGCTGCG- -AAC 425
 DB 348 ATCHDRVASFYCEPHGRTGLLLKXACISNPNCEG- - - - -SNCUTPNVNGKRIC- 398
 QY 426 GCAGTCTGGCCA- - -TCGCCCCGGGCGCAGGGAAGGGCCGCGAGGAGCC- -GCTGA 482
 DB 399 TCRSGTGPACSDQVDECDLGNRCERA- - - - -GKCLNTLGSFECQCLQYGTG 447
 QY 483 GCTTGTGGAGCATGAC- - - - -CAC- - - - -TGCTACTC- - - - -C- - 514
 DB 448 GCEIDVNECISNPQNDATCLDQIGFQCICMPGIEGYVCEINDECAASPCLNHGCM 507
 QY 515 - - - - -CGCTGCGG- - - - -GAACTGTGTACCCG- -GAGTCCCGAGAGGAC- - - 552
 DB 508 XIHFQCOCPKGFNGHLCOYDVDECASTPCKNAGAKLDGNTVTCVTEGYTGHCEVDI 567
 QY 553 - - - - -TC- - - - -ACCTAGCAGGTGGAATCTCTACAGCC- - -GTCA- - - - -TC 589
 DB 568 DECDPDPCHYSGKDGVAFTCLCPQGYTHHCETNINEHSQPCRHGGTCDQRNSYL 627
 QY 590 GACTACATCTCGACTCGAG- -TAGTC- - - - -CTGGCCGAGC- - - - -CAGCC 631
 DB 628 -LCLKGTGPNCENLDDCANPCDSGTCLXKIDGYECACBPYGTGSMCNVWIDECAGSP 686
 QY 632 CTTGACCCCTGATGGCC- - - - -CCACTTCCCATCCAGACA- -GCGAGCTC- - - 679
 DB 687 CHNGGTCEBDGIAGT- - -CRCEGYHDPDTCLESEVNECSNPCHGACRDGLNGYKCDAPG 744
 QY 680 - -GCTC- - - - -CGGAATCT- -GTCA- - - - -TCTC- - - - -CAACG 704
 DB 745 WSGTWCDINNCEBSPNCPNGGTCKDMTSGVICTCRGFSGNQTNINECASPNCLNQ 804
 QY 705 ACAAAGGAGCT- - - - -TTTGCACTGACTCGCGGCTGCTCTG- - - - -ACACTCCAGA 753
 DB 805 TCIDDVAGYKNCPLPYTGATCEVVLAPCATSPCKNSGVCKESDEYFSCVCTGQGG 864

RESULT 13

NTC3_HUMAN
 ID NTC3_HUMAN STANDARD; PRT; 2321 AA.
 AC Q9UM47; Q9UEB3; Q9UPL3; Q9Y6L8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP MEDLINE=97032728, PubMed=8878478;
 RA Joutel A., Corpechot C., Ducros A., Vahedi X., Chabrier H., Mouton P.,
 RA Alauwitsch S., Domenga V., Cecillon M., Marechal E., Maciazek J.,
 RA Vaysiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
 RA Bach J.-F., Bousier M.-G., Tournier-Lasserre E.;
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 RT causing stroke and dementia.";
 RL Nature 383:707-710 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gunel M., Artavanis-Tsakonas S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,

| | | | | | |
|--|---------|---|------|--------------------------|--|
| FT | TRANSEM | 1644 | 1664 | POTENTIAL. | |
| FT | DOMAIN | 1665 | 2321 | CYTOPLASMIC (POTENTIAL). | |
| FT | DOMAIN | 40 | 77 | EGF-LIKE 1. | |
| FT | DOMAIN | 78 | 118 | EGF-LIKE 2. | |
| FT | DOMAIN | 119 | 156 | EGF-LIKE 3. | |
| FT | DOMAIN | 158 | 195 | EGF-LIKE 4. | |
| FT | DOMAIN | 197 | 234 | EGF-LIKE 5. | |
| FT | DOMAIN | 236 | 272 | EGF-LIKE 6. | |
| FT | DOMAIN | 274 | 312 | EGF-LIKE 7. | |
| FT | DOMAIN | 314 | 350 | EGF-LIKE 8. | |
| FT | DOMAIN | 351 | 389 | EGF-LIKE 9. | |
| FT | DOMAIN | 391 | 429 | EGF-LIKE 10. | |
| FT | DOMAIN | 431 | 467 | EGF-LIKE 11. | |
| FT | DOMAIN | 469 | 505 | EGF-LIKE 12. | |
| FT | DOMAIN | 507 | 543 | EGF-LIKE 13. | |
| FT | DOMAIN | 545 | 580 | EGF-LIKE 14. | |
| FT | DOMAIN | 582 | 618 | EGF-LIKE 15. | |
| FT | DOMAIN | 620 | 655 | EGF-LIKE 16. | |
| FT | DOMAIN | 657 | 693 | EGF-LIKE 17. | |
| FT | DOMAIN | 695 | 730 | EGF-LIKE 18. | |
| FT | DOMAIN | 734 | 770 | EGF-LIKE 19. | |
| FT | DOMAIN | 771 | 808 | EGF-LIKE 20. | |
| FT | DOMAIN | 810 | 847 | EGF-LIKE 21. | |
| FT | DOMAIN | 849 | 885 | EGF-LIKE 22. | |
| FT | DOMAIN | 887 | 922 | EGF-LIKE 23. | |
| Query Match 8.7%; Score 656; DB 1; Length 2321; | | | | | |
| Best Local Similarity 23.7%; Pred. No. 1.2e-26; | | | | | |
| Matches 344; Conservative 40; Mismatches 566; Indels 502; Gaps 80; | | | | | |
| QY | 158 | GAGGAAGCCTTTGCAATTAAAGCGGCTG-TGAAGCCAGCGCGGGGGGGGGGGG 216 | | | |
| DB | 35 | GPAAAPPLDGP-CAN-----GRCQLPFRSREAAACLPFGWVERQLEBDPCHSP 86 | | | |
| QY | 217 | CCGAGGGGGCCATTGAATAAGAGCGTGCCTCCAGGAGCGGTCTATAAGTGACCG 276 | | | |
| DB | 87 | CAGRGVQSSVAGTARFSCRCRPGFPGDCLSLPDLCLSSPCAGARCSVGPDRFLC-- 144 | | | |
| QY | 277 | CCGCGGCGAGCGTCG-CGGTTTCAGGTACTGTAGCGACTCTTTTGGTTTCTTTC 335 | | | |
| DB | 145 | SCPFGYQGRSCRSDVDECRVGEPCRHGGTCLNT-----FGSPRCQCPAGYTGPLCENPA 198 | | | |
| QY | 336 | TCTTTGGGGCACCTC-TGGACTCACTCCCGACATGAAG-GC-----GCTGAGCCCGGT- 387 | | | |
| DB | 199 | V-----PCAPSPRNGTQRQSDLTQACALPFGEGQCNVVDVDFGHRCLNGTTC 251 | | | |
| QY | 388 | -----CGCG-GGCTG-CT-----ACGAGCGCGTGTGCTGCTGTGCGAAG 426 | | | |
| DB | 252 | VDGVTYNCQCPPEWTGQFCTEDVDECOLOPNACHG-----GTCFNTLGHSCV 301 | | | |
| QY | 427 | CAGTCTGCCA-----TCGCGCGGGC-----CGAGGGAAG-----GGCCCGGC- 465 | | | |
| DB | 302 | CVNGWTGSCSNIDDDCATAVCFHATCHDRVASFYCAPMGKTGLLLCHLDDACVSNPCH 361 | | | |
| QY | 466 | --AGCTGAGGAGCCGCTGAGCTGTGTGAC-----GACATGAACACTG-C-TACTCCCG 516 | | | |
| DB | 362 | EDALCDTNPVNGRAICT--CPPGFTGACDQDVEDCSIGANPEHLGRVNTQGSFLCQ 418 | | | |
| QY | 517 | CCTGCGGAACTGTGTACCG-----GAGTCCCGAGAG--C-----ACTCAGC 557 | | | |
| DB | 419 | CGRGYTPRCEDVNECLSGPORNQATCLDRIGQFTICMAGFTGTVEVDIDECQSSPC 478 | | | |
| QY | 558 | TTAGCCAGGTGAAATCTACAGCGC-----GTCACTGACTACATCTCGACCTG----- 607 | | | |
| DB | 479 | VNGGVCKRVNGFSCCTSGFSGSTCQLDVDECA-----STPCRNGAKCVDPDGYE 530 | | | |
| QY | 608 | --CAGTAGTCTCT-----GGCGGAGCGAGCCCT----- 634 | | | |
| DB | 531 | CRCAEGFGLCDRNVDCSPDPCHHGRCVDFGASFCACAPGYGTGRCSQVDECRSQP 590 | | | |
| QY | 635 | -----GGAC-----CCCTCATGCGC-----CCACCT-----TCCCATCCA 666 | | | |
| DB | 591 | CRHGGKLLVDKYLRCFSGTGTGNCVFNIDDCASNFTFCVCRDGINRYDCVCPGFT 650 | | | |

| | | | | | |
|------------------------------------|---|--|--|--|--|
| QY | 667 | G-----ACAGC-CGAGCTC-----CCTC-----CGAACTGTGTC- 694 | | | |
| DB | 651 | GPLCNVEINECASPCGEGSCVDGNGFRCLCPGSLPPLCLPPSPHPCAHFPCSHGICY 710 | | | |
| QY | 695 | -----ATCTC-----CAACGACAAAAGAGCTTTTGC-----CA 723 | | | |
| DB | 711 | DAPGFRVCPEPWSGPRCSQSLARDACESQPCRAGTCSDDGNGFFHCTCPFGVQGRQCE 770 | | | |
| QY | 724 | CTGACTCGGCGTGTG-CTGACACCTC-----CAG-AAAG-----CA 758 | | | |
| DB | 771 | LLSPTNPNCEHGRGRCESAFQGLPVCSPQGWQPCQDQVDECAAGPAPCPHGICNTLA 830 | | | |
| QY | 759 | GGTGTCT--GGCGCCCTTTTGTG-----CTGGAGCCCGGGAACCTCTCTCTG-----CCGG 806 | | | |
| DB | 831 | GSFSTCHGGYTGSPSCDQINDCPNCLNGGS---CQDVGVSFSCSLFPGFAGRCARD 887 | | | |
| QY | 807 | AAGCGGAGC-----GCAGGATG-GC-----CCCAACTTCGCCCTGTC--C 845 | | | |
| DB | 888 | VDELSNPCCGPGTCTDHVAFCTCTCPGYGFGHCEQDLPCDSPSSCSFNGGTCVDGVNSFS 947 | | | |
| QY | 846 | CACITGACTTACCAAAATCCCTTCTGTGAGACTAACTGTGTCTCA---GGAGCGAA-- 900 | | | |
| DB | 948 | CLCRPG-YTGAHCQHEADPCLSLRPLHG-GVCSAA---HFGFRCTCLLESFTGPOCOTLVD 1002 | | | |
| QY | 901 | -----GGACTGTGAACCTGTAGCTGAAGAGC-----CAGAGCTAG-----CTC 939 | | | |
| DB | 1003 | WCSRPQPCQNGRCVQTGAYCL-----CPPGWSGRCLDIRSLPCEAAAQIGVRLQOLCQA 1057 | | | |
| QY | 940 | TGGC-----CACCAGCTGGC-----GAC-----GTCACCTGCTCCC----- 972 | | | |
| DB | 1058 | GQCVDEDSHYVCPEGRTGSHCEQVDPCLAQPCQHGCTCRGYMGSGYMCCLPGYNGD 1117 | | | |
| QY | 973 | -----ACCCAC-----CCCCAACTTCTAGGTCTTTTTCAGAGCGTGGAG 1012 | | | |
| DB | 1118 | NCEDVDVDCASQPCQHGSCIDLVARYLCSPPGT---LGVLEINE---DDCGPFPPL 1170 | | | |
| QY | 1013 | GTGTGGAAGAGTGGCTCTCTCCAAACTATGC-----CAAGGC-----GG 1053 | | | |
| DB | 1171 | DSGPRCLHNGTCVDLVGGFRCT-CPGYTGLRCEADINECRSGACHAAHTRDCLQDPGGG 1229 | | | |
| QY | 1054 | CGCGAGAGCTGTCTT-----CTGGTC-----TCCTTGGAGAAAGTTCTGTGTCGCC- 1100 | | | |
| DB | 1230 | FRCLCHAGFSGRPCOTVLSPCESQPCQHGCGCRPSPGFG---GGTFT---CHCAQFPW 1282 | | | |
| QY | 1101 | -----TGATTTATGAATC-----TATAATAGAGTATATA- 1130 | | | |
| DB | 1283 | GPRCERVARSCELOCPVPCQOTPRGRCACPPGLSGPSCRSFPGSPFGASNASCAA 1342 | | | |
| QY | 1131 | -----GGTTTG-----TACCTTTTACAGGAAGGTGACTTTCTGTAA----- 1169 | | | |
| DB | 1343 | PCLHGGSCRPAPLAPFFRCACAGQGTGPRCEAPAAAPVSEBPRCPRAACQAKRGDQRCD 1402 | | | |
| QY | 1170 | --CAATGCGATG 1179 | | | |
| DB | 1403 | RECNSPGCGWDG 1414 | | | |
| RESULT 14 | | | | | |
| BAR3_CHITE STANDARD; PRT; 1700 AA. | | | | | |
| ID | BAR3_CHITE | Q03376; | | | |
| AC | Q03376; | | | | |
| DT | 01-OCT-1993 (Rel. 27, Created) | | | | |
| DT | 01-OCT-1993 (Rel. 27, Last sequence update) | | | | |
| DT | 15-MAR-2004 (Rel. 43, Last annotation update) | | | | |
| DE | Balbani ring protein 3 precursor. | | | | |
| GN | BR3. | | | | |
| OS | Chironomus tentans (Midge). | | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | | | |
| OC | Neoptera; Endopterygota; Diptera; Nematocera; Chironomidae; | | | | |
| OC | Chironomidae; Chironominae; Chironomus. | | | | |
| OX | NCBI_TaxID=7153; | | | | |
| RN | [1] | | | | |

SEQUENCE FROM N.A.
TISSUE=Salivary gland;
MEDLINE=90172404; PubMed=1689777;
Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
"The Balbiani ring 3 gene in Chironomus tentans has a diverged
repetitive structure split by many introns.";
J. Mol. Biol. 211:331-349(1990).
CC -!- FUNCTION: Used by the larvae to construct a supramolecular
CC structure, the larval tube. Balbiani ring protein 3 could play a
CC role as a transport protein that binds to other proteins
CC intracellularly and in the gland lumen in order to prevent these
CC from forming water-insoluble fibers too early.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Salivary gland.
CC -!- DOMAIN: Has 82 approximate repeats of Cys-x-Cys-x-cys.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52263; CAA36506.1; -
DR PIR; S08167; S08167.
DR HSSP; P15358; 1SKZ.
DR InterPro; IPR004153; CXXC-repeat.
DR Pfam; PF03128; CXXC; 71.
KW Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.
SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;

Query Match 8.7%; Score 654.5; DB 1; Length 1700;
Best Local Similarity 22.3%; Pred. No. 1.1e-26;
Matches 351; Conservative 30; Mismatches 650; Indels 545; Gaps 89;

QY 50 GAGTGA-TTTTAAATCTTGTGCGGGA-
DB 94 GCGAGRYWCNQCSCKSTPMSAGCSQIWEKSCACVCPNADKTAPOWNNKDTCCC 153
QY 85 GC-----CTCTC-----CCCGTATCAG-----102
DB 154 GCPVNMQEPADGCTKPLWDVDCCEPLKDDCGKNDWSDSSCSCEKDGKCGQSKI 213
QY 103 -----CGCTTC-----CTCA-----TTCTTTGAA-----TC-CG 125
DB 214 WKNKNCRCICPTAPBAGGCSAPLKWDDDKSCACCPAKMBEKEKCVESGKTWNPTCEG 273
QY 126 CGGCTC-----CGCGTCTTGGGTCTAGACGACCGCGAGGAGCGCTGTTGC 173
DB 274 CAQLNCPDNKANKETCCCEKEVKKNGGVFKDSCSCVCPGGDXKT-----323
QY 174 AATTAAGCGGGCTG-----TGAACGCCC-----AGGGC-----CGCGGGGG 211
DB 324 -----CTAPQVYDVACSCPCPVNMQKPADGCPRPQKWDKEECRC-----363
QY 212 CGGGCGGAGGGGG-----CCATTTTGAATAAGAGCGGTGCTTC-----CA- 255
DB 364 -----ECPVKHKDCKNGKVDLTCQICPRDAPVCTAGKERGESCECKINREKCGAK 419
QY 256 -----GGCAGG-----CTCTATAAGTGA-----CGGC-C 278
DB 420 PLVNMNENTCKVCPADKQMSPGGSGSKSPNKLTCQCECDQSGASKGLKRWADTKCEC 479
QY 279 GCGGCGAGCG-----TGCGCGCG-TTGCAG-----GTCAGTGTAGCGGACTTCTTTG 325
DB 480 QGMPPEGCGKQWISDKCECSPTITCOAPQILDLTNLTCEKCPVNNLAQKEKCKSPRQ 539
QY 326 GTTTTCTTCTCTTTGGG-----GAC-CTCTGGACTC-----ACTCCC-----363
DB 540 WTDKSC--LCESTTPATCEGKQWCGEACQICFSGDKNGCNKXKFPDKPSCCKCKNXP 597

QY 364 -CAGCATGAGGCGC-----TGAGCCCGGTG-----CGCGGTCTGTACGAGCGGTGT 410
DB 598 TCTSPQVWDADDCECKCPKDKQKPGGCGDGGQWNRVCS-----GCPVPRP 645
QY 411 GCTGCTCTCGGAAGAGAGTCTGCG-----CATCGCCCGG-----GCGAGGAGG 458
DB 646 DCTNGQIYININTCAGC-GIDKPSCPKQIYINNTCTDCECPNGMKVEPVGCGAKTW-LDD 703
QY 459 GCCC-----GGCAGCTGAGGAGC-CGCTGAGCTTGTCTGG-----AGC-ACATGAAC 502
DB 704 EQCDDCVPGKPGGCTGAQKWCMDTKCKCKEMPTGCGENNKWCDTETDCVCPQNTC 763
QY 503 CACTGTACTCCCGCTCGGGAAGTACTACCGGAGTCCGAGAGC--ACTCAGTTA 560
DB 764 TAPKVDARTCSICVNPVKNSPVILKDTCCCGCQNVKSKAPKFIENICDA-CPNK 822
QY 561 GCCAGGTGGAATCTACAGCGCTCATCGACTA-----CATTTCTCA-----CCTG--- 607
DB 823 KQCKAPLWSDPFC-----DCVCPNSASMTCLSPKEMNKVTCTCDNPPKPCDCTQK 877
QY 608 -----CAGTAGTCTCTGG-----CGAGCCAGCCCTCG-----ACCCCTGATGG 648
DB 878 WMDKCKGCPNAOTDCAAGKFNDFTCSCGCPSGKLDCTGNTKWSAETCTCGGDNVRN 937
QY 649 C-----CCCCAC-----C-----TTCCATCCAGACAGC-----CGAGC 677
DB 938 CGNLKNFNDNLQCECKKNQEMANCKSPRTWNYDTCKVCKNADSDDDCVKQIWLDDQC 997
QY 678 TCGCTCCGAACTTGTGTC--ATCTCBAACGACAAAGAGCTTTTGCA--CTGACTCGG 732
DB 998 KCGCPASQMTCPANKRPIKSCSC--ECKSPWSPITPOGKKWNEDKCVVECANVTC-- 1054
QY 733 CGGTGTCTCTGACAC-C-----TCCAG-----AACGCAAGTGTCTG-----CGCCCG 772
DB 1055 --GPQRWCDNKCKICPQVNTKCSDKQKFTESKCECETOTQCKDGRFWSNLECGCLD 1112
QY 773 -----TTCTGCTGGGACCCCGGGA-AC-----CTCTC-----CTGCCGAA 808
DB 1113 DKCKPGKQVFNKTCQCKCPNQKPGDTCGNGKDFCLDCKCKKNKPNPANGCTGVQWNE 1172
QY 809 GCGGAGC-----GCGAGGATGGGCCCCCACTTTCGCCC-----TGCCCACTTGACTT 855
DB 1173 BKQCECPKPKKQCPGGQDNNHQCQCGCTPAPTCSNNQKYSNVSCGCGNPKPKN 1232
QY 856 CACCAATCCCTTC-CTGGAGAC-----TAAACCT---GGTGTGAGGAGCGAAGACT 905
DB 1233 -GCPNQIWCNDTCTC-----VCPKNMEKPADNCKTKWMDNEMCCQCKPFC--PEGGCK 1284
QY 906 GTGAATTTGTAGCTGAGAGCAGCAGC-----TAGCTC-----TGCC-----943
DB 1285 GV-----MKWNANTCCECPADKAPASGDKKSNDDSCQCKSKWPCGCGPPNQWNE 1340
QY 944 -----CACCA--GCTGGGCGAGCTCAACC--TG-CTCCACCCACCCCGAGTTCTAAGG 994
DB 1341 KDECKCKSATGNCPAGQTNWNSQTCQCPATGRTGAQVWCVKACCKVCFAQKKCDSP-K 1399
QY 995 TCTTTTTCAGAGCTGGAGGTGTGAAGAGTGGTGTCTCTCCAACTATGCC-----1046
DB 1400 TWDENSCS---CQCPKNRPTGCGNAGRTWDDAT-----CSEKCAATPKCDSPKVDP 1450
QY 1047 AAGCGCGCGGAGAGCTGCT-----C-TTCTGCTCTCTGGAGAAAGTTCTGTGCCC 1100
DB 1451 TTCGC-KCGNPKNLDAGRWDBOKKMTCDLPAIPCHY--EGQVDSNTCK-----CGC 1503
QY 1101 TGAATTTATG-----AACTCTATAATAGATATAGTTTGTACCTTTTACAGGAA 1154
DB 1504 PKETCKQGFSPKSGCKILECNKKDPGCGAKKIWCQETCKCECASEPFRMGCPNRY 1563
QY 1155 GGTGACTTCTGTAAAC 1170
DB 1564 WDXEDCAQCCLTKKVC 1579

RESULT 15

NTC3_RAT

ID NTC3_RAT STANDARD; PRT: 2319 AA.

AC C9R172;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).

GN NOTCH3.

OS Notch signaling (Rat).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OC NCBI_TaxID:10116;

[1]

RN SEQUENCE FROM N.A.

RP Haritunians T., Boulter J., Weinmaster G., Schanen N.C.;

RA "Rattus norvegicus" mRNA for Notch 3.;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

[2]

RN FUNCTION.

RP MEDLINE=21094508; PubMed=11182080;

RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,

RA Honjo T.;

RA "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent neural progenitor cells to an astroglial fate.;"

RT Neuron 29:45-55(2001).

RL [3]

RN TISSUE SPECIFICITY.

RP MEDLINE=21331789; PubMed=11439922;

RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;

RA "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple functional roles for the Notch-DSL signaling system during brain development.;"

RT J. Comp. Neurol. 436:167-181(2001).

RL [4]

CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta1 to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs (By similarity). Acts instructively to control the cell fate determination of CNS multipotent progenitor cells, resulting in astroglial induction and neuron/oligodendrocyte suppression.

CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).

CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.

CC -!- TISSUE SPECIFICITY: Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones.

CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(TW) and a N-terminal fragment N(EC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).

CC -!- PTM: Phosphorylated (By similarity).

CC -!- SIMILARITY: Belongs to the NOTCH family.

CC -!- SIMILARITY: Contains 34 EGF-like domains.

CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.

CC -!- SIMILARITY: Contains 5 ANK repeats.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation


```

Qy 963 -CCC-----TGCTC-----CCACCCAC-----C-----CCCAAGTTCTAAGTCTT 998
Db 1192 RCNCPGYTGLHCEADINECREPGTCHAAHTRDCLQDPGHERCICLPFT-----GPRCQT 1247
Qy 999 T--TCAGAGCCTGGAGGTGTGAAGAGTGGCT---GCTCTCCAACTATGCCAAGCG 1052
Db 1248 ALFPCEQPQHGQCRPSLGRGGLTFTCHCVQFPWGLRCERVARSORELQC----- 1300
Qy 1053 GCGGCAGAGCTGGTCTTCTGTCTCTCT-TGGAGAAAGTTCTGTGCCCTGATTATGAA 1111
Db 1301 ----PVGIFCQ----QTARGPRCACPPGLSGPSCRVSRASPSGATNTSC-AATPCLHGS 1351
Qy 1112 C-----TCTATAATAGATATATAGTTTGTACCTTTTACAGGAAGTGACTT 1162
Db 1352 CLPVQSVPPFRVCAPGCGPRCTPSAPEVPEPRC-----PRAACQ--AKRGDQNCDR 1405
Qy 1163 TCTGTAAACATCGATG 1179
Db 1406 ECN-----SPGCGWDG 1416

```

Search completed: May 7, 2004, 15:08:52
Job time : 35 secs